

VSEBINSKI, PODATKOVNI IN PROCESNI MODEL IZVEDBE ANALIZ PODATKOV

ELENA ČOKOR,^{1,2} ANJA ŠKETA,^{2,3} DRAGO BOKAL^{1,2,3}

¹ Univerza v Mariboru, Fakulteta za naravoslovje in matematiko, Maribor, Slovenija
elena.cokor@student.um.si, drago.bokal@um.si

² DataBitLab d.o.o., Maribor, Slovenija
elena.cokor@student.um.si, anja.sketa@imfm.si, drago.bokal@um.si

³ Inštitut za matematiko, fiziko in mehaniko, Ljubljana, Slovenija
anja.sketa@imfm.si, drago.bokal@um.si

V prispevku predstavljamo metodologijo za standardizirano izvajanje analiz podatkov, ki podjetjem omogoča bolj zanesljivo, pregledno in učinkovito analitično delo. Pristop temelji na enotni vsebinski zasnovi, podatkovnem modelu ter jasno strukturiranem procesu, ki analitikom zagotavlja doslednost in sledljivost, uporabnikom pa boljše razumevanje rezultatov. Z opredelitvijo ključnih podatkovnih elementov in njihovih povezav podatkovni model omogoča, da so analize zgrajene na preverjenih, primerljivih in ustrezno dokumentiranih informacijah. Poseben poudarek namenjamo vprašanju kakovosti podatkov, ki pomembno oblikuje zanesljivost analitičnih zaključkov in odločitev, ki izhajajo iz njih. Metodologija vključuje celoten potek analiz podatkov, od oblikovanja zahtevka in izbire podatkovnih virov do preverjanja kakovosti, priprave algoritma, ovrednotenja rezultatov ter končne ugotovitve. S takšnim pristopom organizacija pridobi robusten analitični okvir, ki podpira razvoj naprednih podatkovnih rešitev ter izboljšuje strateško in operativno odločanje.

DOI

[https://doi.org/
10.18690/um.fov.3.2026.10](https://doi.org/10.18690/um.fov.3.2026.10)

ISBN

978-961-299-124-1

Ključne besede:

analiza podatkov,
proces izdelave analiz,
podatkovni model,
podatkovna zanesljivost,
analitična metodologija,
poslovna analitika



Univerzitetna založba
Univerze v Mariboru

DOI
[https://doi.org/
10.18690/um.fov.3.2026.10](https://doi.org/10.18690/um.fov.3.2026.10)

ISBN
978-961-299-124-1

CONTENT, DATA, AND PROCESS MODELS FOR PERFORMING DATA ANALYSES

ELENA ČOKOR,^{1,2} ANJA ŠKETA,^{2,3} DRAGO BOKAL^{1,2,3}

¹ University of Maribor, Faculty of Natural Sciences and Mathematics, Maribor, Slovenia
elena.cokor@student.um.si, drago.bokal@um.si

² DataBitLab d.o.o., Maribor, Slovenia

elena.cokor@student.um.si, anja.sketa@imfm.si, drago.bokal@um.si

³ Institute of Mathematics, Physics and Mechanics, Maribor, Slovenia
anja.sketa@imfm.si, drago.bokal@um.si

Keywords:

data analysis,
data analysis development
process,
data model, data reliability,
analytical methodology,
business analytics

This paper presents a methodology for the standardized execution of data analyses, enabling organizations to perform more reliable, transparent, and efficient analytical work. The approach is based on a unified content framework, a data model, and a clearly structured process that ensures consistency and traceability for analysts while improving end users' understanding of the results. By defining key data elements and their relationships, the data model ensures that analyses are built on verified, comparable, and properly documented information. Special emphasis is placed on data quality, which critically shapes the reliability of analytical conclusions and the decisions derived from them. The methodology covers the entire data analysis lifecycle—from request formulation and selection of data sources to data quality verification, algorithm preparation, result validation, and final presentation of findings. Such an approach provides organizations with a robust analytical framework that supports the development of advanced data solutions and enhances both strategic and operational decision-making.



University of Maribor Press

1 Uvod

V podjetjih se dnevno ustvarja ogromna količina podatkov, od meritev v poslovnih procesih do informacij o sredstvih in obratovalnih pogojih (Davenport, 2006). Kljub bogatim virom ostaja ključni izziv, kako iz podatkov pravočasno in zanesljivo pridobiti informacije za podporo odločanju. Odgovor na ta izziv je strukturirana analiza podatkov, ki zagotavlja, da so podatki preverjeni, razumljeni in pripravljeni za obdelavo, kar jim daje dodatno vrednost. Tak pristop ni le izvedba izračunov, temveč standardiziran način razmišljanja o analizah: od njihovega oblikovanja in izbire podatkov do preverjanja kakovosti in razumljive predstavitve rezultatov. Prispevek temelji na študiji (Bokal idr., 2020; Lustig idr., 2010), ki povezuje vsebinski, podatkovni in procesni vidik analitičnega dela.

S strukturiranim pristopom se rešuje pogosta težava, ko se isti problem v podjetju rešuje na več načinov z različnimi rezultati. Standardizirana analiza uvaja strukturo za ponovljiv proces dela, uporabnikom pa zagotavlja zanesljive in sledljive rezultate.

V prispevku obravnavamo celovit model, ki se uporablja pri pripravi analiz in prikazov. Pokažemo, kako jasna opredelitev vsebine in enotna struktura postopka vodita do natančnejših in doslednejših rezultatov. Na konkretnem primeru iz prakse prikažemo, kako takšen pristop poenostavi pripravo analiz, zmanjša možnost napačnih interpretacij in izboljša učinkovitost odločanja v podjetju.

2 Teoretične osnove standardiziranega pristopa k podatkovni analitiki

V podjetjih se podatki zbirajo iz številnih različnih virov, kar ustvarja okolje z veliko količino informacij, a neenakomerno kakovostjo. Ker je od kakovosti odvisna učinkovitost uporabe podatkov v praksi, je ključno zagotoviti njihovo enotno obravnavo, dosledno preverjanje in zanesljivo interpretacijo pri odločanju. Strukturirana analiza podatkov je nastala kot odgovor na ta izziv. Njen namen je vzpostaviti standardiziran način razmišljanja o analitičnih postopkih, ki omogoča preglednost, ponovljivost in skladnost z vsebinskimi cilji. S tem presega klasično analitiko, saj ustvarja celovit okvir, v katerem je vnaprej jasno, kaj analiziramo, zakaj je to pomembno in kako zagotoviti primernost podatkov za obdelavo.

2.1 Strukturirani pristop

Strukturirana analiza podatkov je celovit proces, ki povezuje vsebinski, podatkovni in procesni vidik dela. Vsebinski del določa smisel analize z vidika uporabnika, podatkovni del se ukvarja z razpoložljivostjo in kakovostjo podatkov, procesni del pa določa zaporedje korakov izvedbe. Takšen pristop zagotavlja enotno strukturo analiz ne glede na njihovo vsebino, kar omogoča primerljivost rezultatov, njihovo integracijo v širše podatkovno okolje podjetja ter zmanjšuje tveganje napačnih interpretacij (Rizzi, 2009).

Analitični postopek poteka kot jasno opredeljeno zaporedje dejavnosti. Najprej se opredeli zahteva uporabnika in opiše problem, nato pa se oblikujejo cilji in vsebinska vprašanja. Sledi preverjanje kakovosti podatkov (popolnost, pravilnost, časovna ustreznost in konsistentnost), saj je kakovost izračuna neposredno odvisna od vhodnih podatkov. Po preverjanju analitik določi logiko obdelave, ki vključuje izračune in filtracije s sprotim preverjanjem smiselnosti. Zaključni del predstavlja priprava prikazov, prilagojenih uporabniku. Ključna prednost procesa je visoka sledljivost, saj je vsak prikaz mogoče povezati z vhodnimi podatki in postopkom obdelave (Mahnic, 2013).

Kljub prednostim ima model določene omejitve. Metodologija ne more nadomestiti manjkajočih informacij, preverjanje podatkov pa je lahko pri velikih naborih časovno zahtevno. Proces je odvisen od ažurnosti virov, učinkovitost pristopa pa od organizacijske zrelosti in pripravljenosti zaposlenih na standardizirani način dela, saj celovitost metodologije pride do izraza le, če jo dosledno uporabljajo analitiki in tudi naročniki analiz.

2.2 Obstoječi standardi in tehnološke rešitve

V praksi obstajajo številna orodja in programske rešitve, ki podpirajo posamezne faze analize podatkov, kot so zbiranje in shranjevanje podatkov, njihova obdelava, upodobitev ter poročanje rezultatov. Takšna orodja analitikom omogočajo učinkovitejšo izvedbo izračunov in pripravo prikazov, vendar se večinoma osredotočajo na tehnično izvedbo analiz ali na podporo odločanju v obliki končnih poročil in nadzornih plošč.

Poleg konkretnih programskih rešitev, kot sta npr. Great Expectations za preverjanje kakovosti ali Apache Atlas za sledljivost podatkov (Moses idr., 2022), obstajajo tudi

metodološki okviri, ki opisujejo faze analitičnega procesa. Najbolj razširjen med njimi je standard CRISP-DM (Cross-Industry Standard Process for Data Mining)(Shearer, 2000), ki poudarja pomen razumevanja problema, priprave podatkov ter interpretacije rezultatov.

Ti pristopi prispevajo k večji sistematičnosti analitičnega dela, vendar pogosto ne opredeljujejo natančno strukture podatkov in njihove kakovosti v povezavi s posameznimi analizami (Batini & Scannapieco, 2016; Henderson idr., 2024). Skupna značilnost večine obstoječih pristopov je, da obravnavajo vsebinski, podatkovni in procesni vidik analize ločeno ali implicitno, brez enotnega modela, ki bi te vidike jasno povezal. Posledično je sledljivost med vsebinskim namenom analize, uporabljenimi podatki in končnimi prikazi pogosto omejena, kar lahko oteži razumevanje rezultatov in njihovo ponovljivo uporabo v praksi. Predstavljeni pristop v tem prispevku zato ne nadomešča obstoječih orodij, temveč jih dopolnjuje z jasno strukturiranim vsebinskim, podatkovnim in procesnim modelom, ki omogoča, da so analize zasnovane, izvedene in interpretirane na enoten in sledljiv način.

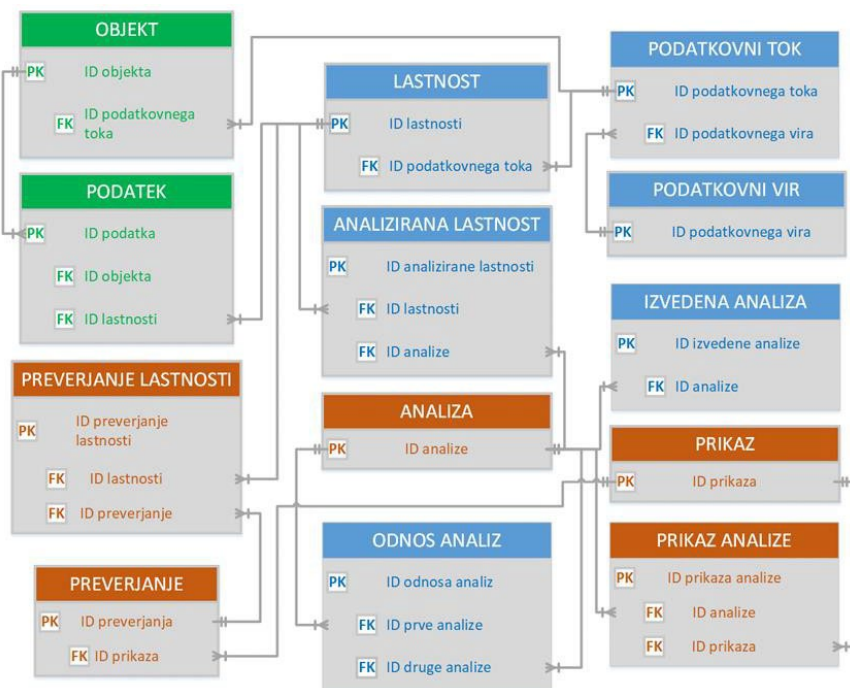
3 Podatkovni in procesni model

Podatkovni in procesni model predstavljata hrbtenico analitičnega delovanja, saj opredeljujeta elemente analize in povezave med njimi. Model analitiku omogoča razumevanje potrebnih podatkov, njihove organizacije, lastnosti in poteka v postopku. Takšen pristop zagotavlja preglednost, enoznačno interpretacijo rezultatov ter razvoj modularnih in nadgradljivih analiz, primernih za vključitev v sistem digitalnega dvojčka elektroenergetskega omrežja.

V središču podatkovnega modela je podatkovni model analize (slika 1), ki določa cilje, pričakovane rezultate, vire podatkov in logiko obdelave. Vsaka analiza je povezana z izvedeno analizo, ki predstavlja dejanski potek obdelave na konkretni podatkovni množici v določenem času. Razlikovanje med obema entitetama omogoča večkratno uporabo iste vsebinske zasnove v različnih obdobjih, kar povečuje ponovljivost in sledljivost.

Model vključuje tudi entiteto prikaz (slika 1), ki zajema način predstavitve rezultatov. Prikaz ni le vizualna komponenta, temveč zaključni element logične verige. Model zagotavlja, da vsak prikaz temelji na jasnih pravilih in je povezan z ustrezno izvedeno analizo, kar omogoča povratno sledljivost nazaj do vhodnih podatkov. Ta povezljivost je ključna za verodostojnost, saj uporabnik natančno vidi, kako je rezultat nastal.

Pomemben del modela so entitete, povezane s podatki in njihovimi lastnostmi (slika 1). Podatek je opredeljen z izvorom, časovno pripadnostjo in kakovostjo (popolnost, rednost, pravilnost, konsistentnost) (Tayi & Ballou, 1998). Model določa način preverjanja teh lastnosti, kar analitiku omogoča razumevanje omejitev virov še pred izvedbo in bistveno zmanjšuje tveganje napačnih rezultatov.



Slika 1: Shema entitetnega modela podatkovne strukture, ki povezuje ključne gradnike analize

Vir: (Bokal idr., 2020)

Procesni model dopolnjuje podatkovni del z natančnim opisom poti podatka od vstopa v analizo do končnega prikaza (Bratuša, 2019). Opredeljuje ključne korake: razumevanje zahteve, oblikovanje zasnove, preverjanje podatkov, priprava algoritma, izvedba in interpretacija. Ta standardizacija omogoča primerljivost dela različnih analitikov in združljivost rezultatov.

Modela skupaj ustvarjata stabilen okvir, ki zmanjšuje možnosti napak in poenostavlja komunikacijo. Analitik se lahko osredotoči na vsebino problema, saj struktura zagotavlja transparentno in sledljivo izvedbo po preverjenih pravilih.

4 Proces izdelave analiz

Proces izdelave analiz je zasnovan tako, da omogoča pregledno, ponovljivo in vsebinsko utemeljeno izvajanje vseh analiz v organizaciji. Temelji na enotni metodologiji, ki povezuje uporabniško zahtevo, podatkovne preveritve in analitično obdelavo v jasno zaporedje korakov (slika 2). S tem se zagotavlja, da je vsaka analiza izvedena na enak način, ne glede na tematiko ali zahtevnost, rezultati pa so vedno sledljivi, primerljivi in pripravljeni za uporabo pri odločanju. Namen procesa ni le tehnična izvedba analize, temveč ustvarjanje okolja, kjer analitik in naročnik delujeta usklajeno, podatki pa so obravnavani tako, da odražajo dejansko stanje in omogočajo zanesljive zaključke.

4.1 Pregled procesa priprave analiz

Priprava analize se začne z razumevanjem vsebinskih ciljev, ki jih postavi naročnik. Analitik najprej oblikuje splošno predstavo o tem, kaj želimo ugotoviti, kateri podatki bodo potrebni in katere omejitve lahko pričakujemo. Ta pregled služi kot orientacija za nadaljnje korake in zagotavlja, da je analiza zasnovana na jasnem problemu. Pregled procesa omogoča, da se že v začetku prepoznajo morebitni izzivi, kot so omejitve podatkov, časovni okvir ali zahtevnost interpretacije.

4.2 Zahtevek

Zahtevek predstavlja formalizirano opredelitev analize. V njem naročnik navede namen analize, ključna vprašanja ter pričakovane rezultate. Dobro oblikovan zahtevek usmerja analitika, saj pojasni, kaj je bistvo analize in zakaj je ta potrebna. V

tej fazi se vzpostavi skupno razumevanje ciljev, kar zmanjša možnost napačnih interpretacij in zagotavlja, da analiza ostane osredotočena na dejanske potrebe uporabnika.

4.3 Vir podatkov

Ko je zahtevek jasen, analitik določi, iz katerih virov bodo podatki črpani. Podatki v podjetjih prihajajo iz številnih sistemov, zato je pomembno razumeti njihovo zanesljivost, izvor in strukturo. Pri izboru vira analitik oceni, ali podatki odražajo realno stanje in ali so dovolj podrobni za namen analize. Ta korak je ključen, saj pravilna izbira vira neposredno vpliva na kakovost nadaljnjih rezultatov.

4.4 Preverjanje kakovosti podatkov

Preverjanje kakovosti podatkov je osrednji element. Analitik pregleda, ali so podatki popolni, pravilni, časovno skladni in notranje konsistentni. Namen preverjanja ni izločanje podatkov, temveč razumevanje njihovih omejitev in vplivov na analizo. S tem se omogoča, da je interpretacija rezultatov realistična in da morebitne pomanjkljivosti ne vodijo do napačnih zaključkov. Kakovost podatkov določa zanesljivost celotne analize, zato je ta korak obravnavan skrbno in sistematično.

4.5 Algoritem analize

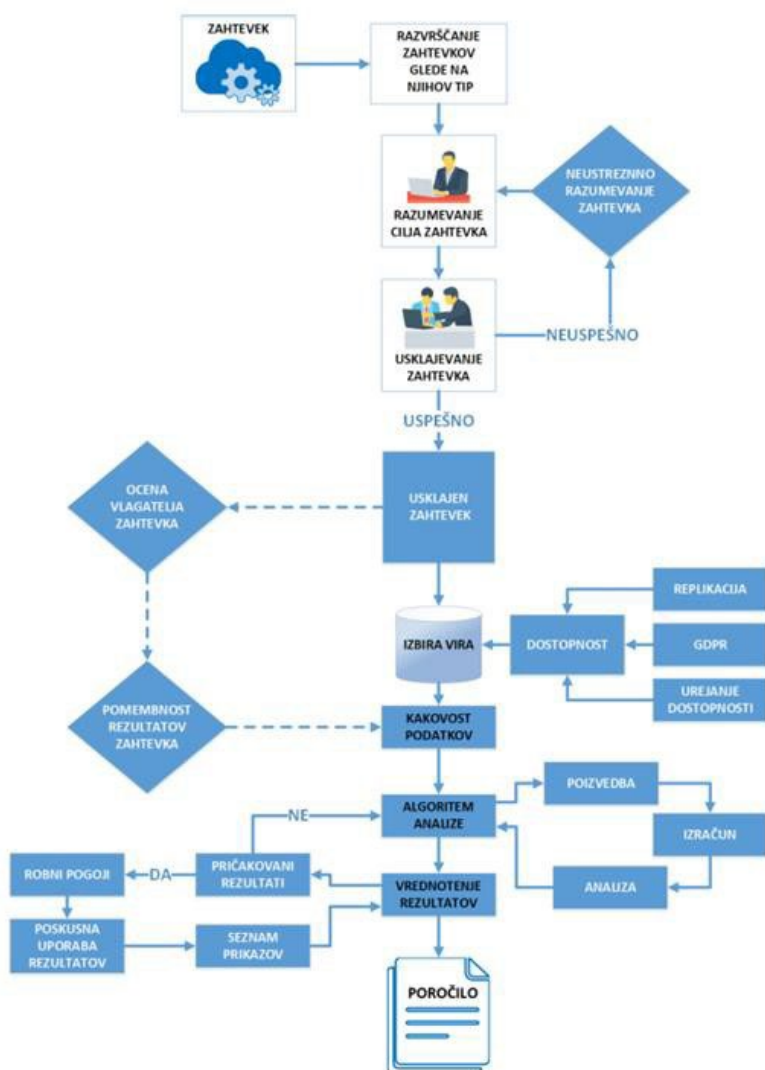
Ko so podatki preverjeni, analitik pripravi logično zasnovo algoritma. Na abstraktni ravni algoritem določa postopke, s katerimi bodo podatki obdelani, merila, ki bodo uporabljena pri vrednotenju, ter pravila, ki vodijo do končnega rezultata. Pri tem je pomembno, da je logika enostavna za razumevanje, sledljiva in ponovljiva. Dobro zasnovan algoritem omogoča dosledne rezultate ter lažje posodobitve, če se analiza v prihodnosti ponavlja ali nadgrajuje.

4.6 Preverjanje rezultatov algoritma

Po izvedbi izračunov analitik preveri, ali rezultati smiselno odražajo stanje, ki ga analiza obravnava. Preverjanje vključuje pregled vmesnih vrednosti, zaznavanje morebitnih anomalij ter oceno, ali so zaključki v skladu s pričakovanji. Namen tega koraka je zagotoviti, da so rezultati ne le pravilno izračunani, temveč tudi vsebinsko

razumljivi in uporabni. Če se pojavijo odstopanja, se analitik vrne korak nazaj in preveri logiko algoritma in tudi kakovost vhodnih podatkov.

4.7 Poročilo analize



Slika 2: Standardiziran proces izdelave analiz podatkov

Vir: (Bokal idr., 2020)

Zaključni rezultat procesa je poročilo, ki združuje vse ključne elemente analize. Poročilo predstavlja potek dela, glavne ugotovitve in interpretacije ter jasno nakaže omejitve, ki vplivajo na končno razumevanje rezultatov. Pripravljeno mora biti tako, da ga razumeta tako analitik in tudi uporabnik, ki ni neposredno sodeloval v postopku. Poročilo ima pomembno vlogo pri zagotavljanju sledljivosti, saj omogoča, da se analiza kadarkoli ponovno preveri ali izvede z enako logiko na novih podatkih.

5 Stopnje in vpliv (ne)kakovostnih podatkov

Kakovost podatkov je eden najpomembnejših dejavnikov, ki določajo zanesljivost in uporabnost analize. V praksi podatki nikoli niso popolni, zato mora pristop analize zagotoviti, da so omejitve podatkov prepoznane, pravilno dokumentirane in smiselno upoštevane pri interpretaciji rezultatov. Za razumevanje zanesljivosti analize je zato ključno, da podatke ne obravnavamo zgolj kot številčne vrednosti, temveč kot elemente z lastnostmi, ki vplivajo na njihovo interpretacijo in analitično vrednost (Batini & Scannapieco, 2016). Te lastnosti tvorijo različne stopnje kakovosti, ki omogočajo, da analitik oceni, kako močno lahko zaupa izračunanim rezultatom.

Podatki z visoko stopnjo kakovosti so popolni, pravilni in časovno skladni, kar pomeni, da jih je mogoče neposredno uporabiti pri izračunih. Rezultati, ki temeljijo na takšnih podatkih, imajo visoko stopnjo verodostojnosti, saj analitik lahko sklepa, da odražajo dejanske razmere. Kadar so podatki delno nepopolni ali vključujejo manjša odstopanja, je treba njihovo kakovost označiti kot srednjo. Takšni podatki še vedno omogočajo izvedbo analize, vendar mora analitik natančno navesti, kako morebitne vrzeli vplivajo na razumevanje rezultatov. Podatki z nizko stopnjo kakovosti so tisti, ki vsebujejo več napak, so nepopolni ali niso dovolj natančni za namen analize. Tudi takšni podatki niso brez vrednosti, vendar zahtevajo previdnost in lahko vodijo do večje stopnje negotovosti.

Vpliv nekovostnih podatkov postane najbolj očiten pri interpretaciji rezultatov. Tudi tehnično pravilno izvedena analiza ne more zagotoviti zanesljivih zaključkov, če temeljni podatki niso ustrezni. Zato je eden ključnih elementov tega pristopa, da se vsak podatek obdela skupaj z oceno kakovosti, ki analitiku omogoča razumeti, kako podatkovne omejitve vplivajo na obseg in natančnost rezultatov. Na ta način se zmanjšuje tveganje napačnih odločitev, ki bi temeljile na preveč optimistični

interpretaciji ali napačnih predpostavkah. Poleg tega ocena kakovosti podatkov uporabniku omogoča realnejše razumevanje rezultatov, saj mu jasno pokaže, kje so meje zanesljivosti analize.

Stopnje kakovosti podatkov imajo pomembno vlogo tudi pri dolgoročni rabi analiz v organizaciji. Tak pristop omogoča, da se podatki ne obravnavajo enkratno, temveč se sčasoma izboljšujejo preko ponavljajočih se procesov preverjanja. Z vsako izvedbo analize podjetje bolje razume, kje prihaja do največjih vrzeli, in lahko usmerja napore v izboljšave. Takšen krožni proces povečuje zanesljivost podatkovnega okolja in omogoča razvoj bolj naprednih analitičnih pristopov. Kakovost podatkov torej ni le pogoj za dobro analizo, temveč tudi gonilo stalnega izboljševanja analitičnega ekosistema.

Razumevanje kakovosti podatkov je zato bistveno za pravilno interpretacijo rezultatov in za dolgoročno zanesljivost analitičnih postopkov. Tak pristop zagotavlja, da je ta vidik vedno vključen v proces, kar omogoča bolj premišljeno uporabo analiz v podporo tehničnim in strateškim odločitvam.

6 Zaključek

Standardizirani pristop k podatkovni analitiki je celovit in zanesljiv okvir za pripravo, izvedbo in predstavitev analiz v podjetjih. Z vzpostavitvijo enotnega vsebinskega, podatkovnega in procesnega modela analize omogoča dosleden način dela, ki analitikom zagotavlja večjo preglednost, ponovljivost in sledljivost, uporabnikom pa večje zaupanje v rezultate. Standardizirani postopki od zahteve do poročila ustvarjajo okolje, kjer vsaka analiza temelji na preverjenih podatkih, jasno opredeljenih ciljih in razumljivo predstavljenih rezultatih. Tak pristop ne izboljšuje le kakovosti posameznih analiz, temveč tudi celotno podatkovno kulturo v podjetju.

Poseben pomen metodologije izhaja iz poudarjanja kakovosti podatkov kot ključnega elementa analitičnega dela. Z upoštevanjem stopenj njihove zanesljivosti in njihovih omejitev se rezultati interpretirajo realistično in odgovorno, kar zmanjšuje tveganje napačnih odločitev. Pristop omogoča, da se sčasoma izboljšuje tudi samo podatkovno okolje, saj sistematično preverjanje podatkov usmerja pozornost na področja, kjer so potrebne izboljšave. Standardizirana analiza tako ne

deluje le kot posamezen analitični postopek, temveč kot dolgoročen mehanizem za dvig kakovosti podatkov in podporo strateškemu razvoju organizacije.

S predstavljenim modelom pripravljenosti podatkov, procesnim potekom analize in metodologijo poročanja takšna analiza prispeva k večji preglednosti, učinkovitosti in strokovnosti delovanja. Rezultati analiz postanejo bolj primerljivi, postopki bolj predvidljivi, sodelovanje med naročniki in analitiki pa bolj usklajeno. V kontekstu nenehno rastoče količine podatkov in vse večjih zahtev po njihovi zanesljivi uporabi se tak pristop izkaže kot nepogrešljiv del analitičnega okolja podjetja.

Viri in literatura

- Batini, C., & Scannapieco, M. (2016). *Data and Information Quality: Dimensions, Principles and Techniques*. Springer International Publishing. <https://doi.org/10.1007/978-3-319-24106-7>
- Bokal, D., Bratuša, A., Goričan, A., & Tertinek, Š. (2020). Priprava vsebinskega, podatkovnega in procesnega modela analiz za DAC (No. 10/2019). DataBitLab d.o.o.
- Bratuša, A. (2019). *Modeli upravljanja s kakovostjo podatkov in simulacije vrednotenja kakovosti podatkov* [Thesis, Univerza v Mariboru, Fakulteta za naravoslovje in matematiko]. <https://dk.um.si/IzpisGradiva.php?id=74337>
- Davenport, T. (2006). Competing on Analytics. *Harvard business review*, 84, 98–107, 134.
- Henderson, D., Earley, S., Bradley, C., & Data Administration Management Association (Ur.). (2024). *DAMA-DMBOK: Data management body of knowledge* (Second edition, revised, first printing). Technics Publications.
- Lustig, I., Dietrich, B., Johnson, C., & Dziekan, C. (2010). The analytics journey. *Analytics Magazine*, 11–13.
- Mahnic, V. (2013). Improving Software Development through Combination of Scrum and Kanban. <https://www.semanticscholar.org/paper/Improving-Software-Development-through-Combination-Mahnic/11bdd5d55a537bc919b7235aec1ad6f7cc6e4c>
- Moses, B., Gavish, L., & Vorwerck, M. (2022). *Data quality fundamentals: A practitioner's guide to building trustworthy data pipelines* (First edition). O'Reilly Media, Inc.
- Rizzi, S. (2009). What-if analysis. V *Encyclopedia of Database Systems* (str. 3525–3529). Springer.
- Shearer, C. (2000). The CRISP-DM model: The new blueprint for data mining. 13–22.
- Tay, G. K., & Ballou, D. P. (1998). Examining data quality. *Communications of the ACM*, 41(2), 54–57. <https://doi.org/10.1145/269012.269021>