ALATI ZA OBRADU PODATAKA

Česte zablude u radu sa podacima

Kako razlikovati veliku tabelu, Big Data, BI analizu i objašnjiv model?

Predavač: dr Dušan Stefanović

01 **Big Data**

02 **BI i analiza**

03 **Komunikacija**

04 **Objašnjivost**

Četiri zablude menjaju način rada

PREGLED TEME

✖ Mnogo redova = Big Data	→	Big Data zavisi od više dimenzija
✖ Svaka poslovna procena = BI	→	BI koristi podatke i merljive pokazatelje
✖ Data Science = programiranje	→	Rezultat mora biti razumljiv i primenljiv
✖ Tačan model = objašnjiv model	→	Složena odluka može biti teško razumljiva

Ispravna definicija menja izbor infrastrukture, metode i poslovne odluke.

Mnogo redova nije isto što i Big Data

ZABLUDA 1

2.000 REDOVA ≠ **BIG DATA**

može biti obična tabela

zahteva procenu celog problema

Ključno pitanje: da li podaci po količini, brzini i raznovrsnosti prevazilaze mogućnosti uobičajenog sistema?

Broj redova je samo jedan signal — ne i definicija.

Big Data opisuje više dimenzija

PET V KARAKTERISTIKA

V	V	V	V	V
VOLUME	**VELOCITY**	**VARIETY**	**VERACITY**	**VARIABILITY**
količina	**brzina**	**raznovrsnost**	**pouzdanost**	**promenljivost**
milijarde zapisa	podaci svake sekunde	tekst, slika, video	šum i netačni podaci	izvori i intenzitet se menjaju

Big Data nastaje kada se više izazova pojavi istovremeno.

Dva skupa iste veličine nisu isti problem

PRIMER INTERNET PRODAVNICE

OBIČAN SKUP

50.000 porudžbina
isti format
jedno ažuriranje dnevno

Može se obraditi na jednom računaru.

BIG DATA PROBLEM

50.000 događaja u sekundi
klikovi + slike + tekst
obrada u realnom vremenu

Zahteva skalabilnu i distribuiranu obradu.

Složenost zavisi od ponašanja podataka, ne samo od njihovog broja.

Tri pitanja pre oznake Big Data

PRAKTIČNA PROVERA

1. **Može li jedan sistem pouzdano da skladišti i obradi podatke?**
2. **Koliko brzo podaci nastaju i koliko brzo rezultat mora biti spreman?**
3. **Da li podaci dolaze iz mnogo izvora i u različitim formatima?**

Ako uobičajeni pristup radi pouzdano, Big Data tehnologija možda nije potrebna.

BI pretvara istoriju u odluku

POSLOVNA INTELIGENCIJA

1 **PODACI**

prodaja, zalihe, kupci

→

2 **BI ANALIZA**

šta, kada i gde

→

3 **UVID**

zašto se rezultat promenio

→

4 **AKCIJA**

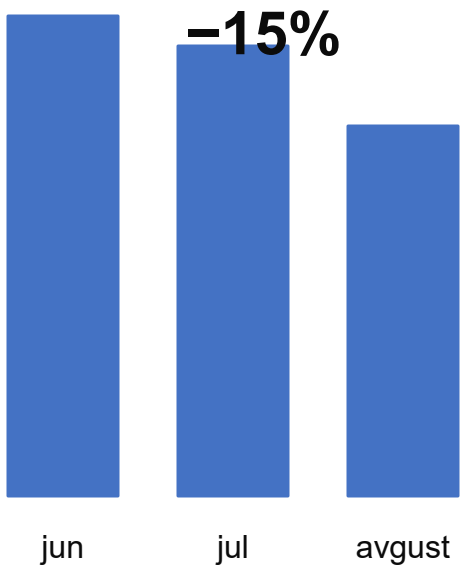
promena plana ili procesa

BI ne završava grafikonom — cilj je bolja poslovna odluka.

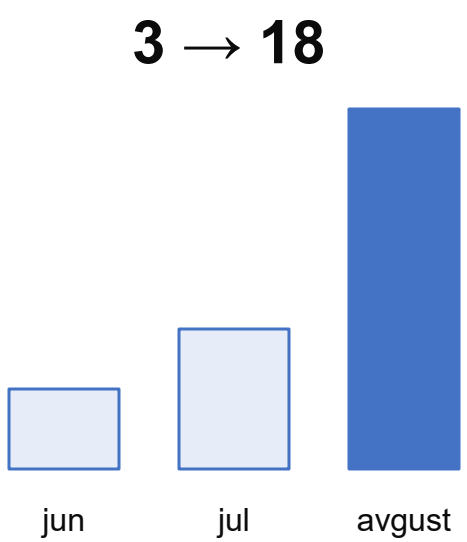
Pad prodaje postaje merljiv problem

BI PRIMER

PRODAJA



NEDOSTUPNI PROIZVODI



POSLOVNI UVID

Pad prodaje vremenski se poklapa sa rastom broja proizvoda kojih nema na stanju.

Podatak opisuje promenu; analiza povezuje promenu sa mogućim uzrokom.

Brojevi i objašnjenja imaju različite uloge

DVA PRISTUPA

KVANTITATIVNA ANALIZA

brojevi • mere • statistika

Koliki je problem?
Kada je nastao?

Primer: prodaja je pala 18%.

KVALITATIVNA ANALIZA

komentari • intervjui • iskustvo

Kako korisnici opisuju problem?
Zašto su nezadovoljni?

Primer: kupci navode dugu isporuku.

Brojevi mere obim problema; opisni podaci pomažu da se razume kontekst.

Najbolji zaključak spaja oba pristupa

PRIMER ZADRŽAVANJA KUPACA

PONOVLJENE KUPOVINE −18%

+

„ISPORUKA JE PRESPORA“

↓

ZAKLJUČAK

Pad lojalnosti se poklapa sa nezadovoljstvom brzinom isporuke — potrebno je proveriti logistički proces.

Kombinovani dokaz je korisniji od jedne izolovane metrike.

SWOT i BI nisu ista analiza

VAŽNO RAZGRANIČENJE

SWOT

snage • slabosti
prilike • pretnje

strateška procena

Može uključiti iskustvo i procenu.

BI

prihod • vreme dostave
reklamacije • tržišni udeo

merljivi poslovni pokazatelji

Oslanja se na podatke iz poslovnih sistema.

SWOT postaje uverljiviji kada se njegove tvrdnje potkrepe BI podacima.

Data Science nije samo pisanje koda

ZABLUDA 2

1	2	3	4
KOD	**MATEMATIKA**	**DOMEN**	**KOMUNIKACIJA**
automatizuje obradu	**formalizuje odnose**	**daje poslovni smisao**	**pretvara rezultat u odluku**

Vrednost nastaje tek kada su tehnički rezultat, kontekst i odluka povezani.

Model bez jasnog značenja ostaje tehnički eksperiment.

Grafikon se bira prema pitanju

VIZUELIZACIJA PODATAKA

Kako se vrednost menjala?	→	**LINIJSKI GRAFIKON**	prodaja po mesecima
Koja kategorija je najveća?	→	**STUBIČASTI GRAFIKON**	prodaja po proizvodu
Da li su promenljive povezane?	→	**DIJAGRAM RASIPANJA**	cena i potražnja
Koliki deo pripada kategoriji?	→	**SLOŽENI STUBIĆ**	udeo kanala prodaje

Vizuelizacija je dobra kada odgovor postaje lakše uočljiv.

Alat se bira prema zadatku

OD PROBLEMA KA TEHNOLOGIJI

Brza tabela i jednostavan grafik	→	**Excel**
Statistička analiza kroz interfejs	→	**SPSS / Stata**
Upiti nad relacionom bazom	→	**SQL**
Automatizacija i ML model	→	**Python / R**
Velika distribuirana obrada	→	**Spark / cloud platforma**

Popularnost alata nije kriterijum; kriterijum je pouzdano rešenje problema.

Najjednostavniji dovoljan alat je dobar izbor

DVA PRAKTIČNA SCENARIJA

MESEČNI IZVEŠTAJ

500 redova
jedna osoba
jednokratna analiza

EXCEL

Brže se izrađuje i lakše proverava.

DETEKCIJA PREVARE

milioni transakcija
obrada u realnom vremenu
automatizovan model

SQL + PYTHON + INFRASTRUKTURA

Skalira se, prati i radi bez ručnog pokretanja.

Dobar inženjerski izbor je dovoljno snažan — ne nepotrebno složen.

Tačan model ne mora biti lako objašnjiv

ZABLUDA 3

ULAZI

- prihod
- dug
- istorija
- starost računa

→

DUBOKA NEURONSKA MREŽA

veliki broj parametara
i nelinearnih veza

→

ODLUKA
RIZIK: 82%

Računanje je poznato; ljudsko objašnjenje pojedinačne odluke može biti teško.

Kreditna odluka pokazuje problem crne kutije

PRIMER ODGOVORNOG AI SISTEMA

MODEL

Zahtev za kredit
→ rizik 82%
→ zahtev odbijen

?

BANKA MORA DA RAZUME

1. **koji podaci su bili važni**
2. **da li je odluka dosledna**
3. **da li postoji nepravedna pristrasnost**
4. **kako objasniti rezultat klijentu**

U osetljivim odlukama objašnjivost je deo kvaliteta sistema.

XAI približava odluku čoveku

OBJAŠNJIVA VEŠTAČKA INTELIGENCIJA

MODEL → **OBJAŠNJENJE** → **ČOVEK**

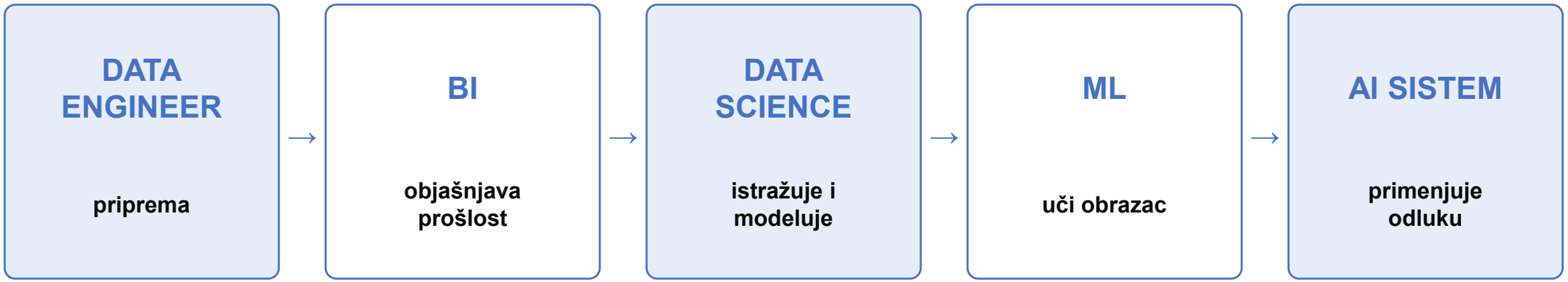
1. glavni faktori odluke
2. stepen pouzdanosti
3. ograničenja objašnjenja

Objašnjenje mora biti razumljivo, smisleno i dovoljno verno ponašanju sistema.

XAI ne uklanja svu složenost, ali omogućava odgovorniju proveru odluka.

Discipline se preklapaju

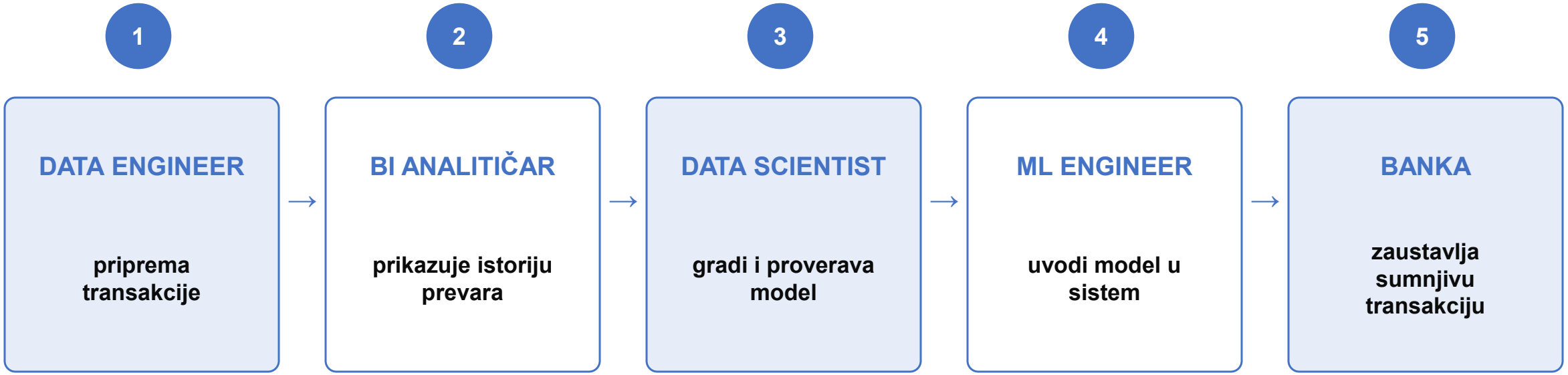
JEDAN KONTINUALNI PROCES



Granice pomažu učenju, ali se realni projekti ne zaustavljaju na granici discipline.

Jedan problem angažuje više uloga

PRIMER BANKARSKE PREVARE



Poslovna vrednost nastaje tek kada rezultati faza rade zajedno.

Pitanje 1

PROVERA ZNANJA

Koji opis najbolje ukazuje na Big Data problem?

A **Tabela sa 2.000 ocena**

B **PDF dokument od 300 strana**

C **Milioni različitih događaja koji stižu u realnom vremenu**

D **Excel tabela sa mnogo kolona**

Izaberite jedan odgovor i obrazložite ga jednom rečenicom.

Pitanje 2

PROVERA ZNANJA

Koja tvrdnja najbolje opisuje BI?

A **BI je svaka procena menadžera**

B **BI povezuje, analizira i vizuelizuje poslovne podatke**

C **BI služi samo za programiranje modela**

D **BI je drugi naziv za duboko učenje**

Razmišljajte o pitanju na koje BI treba da odgovori.

Pitanje 3

PROVERA ZNANJA

Koji par pravilno povezuje pristup i primer?

A **Kvantitativno — komentar kupca**

B **Kvalitativno — prihod u evrima**

C **Kvantitativno — pad prodaje od 18%**

D **Kvalitativno — prosečno vreme isporuke**

Prvo odredite da li je informacija brojčana ili opisna.

Pitanje 4

PROVERA ZNANJA

Zašto se duboki model često naziva „crnom kutijom“?

A Ne znamo koje matematičke operacije izvršava

B Uvek daje slučajan rezultat

C Teško je dati jednostavno objašnjenje pojedinačne odluke

D Ne koristi podatke tokom obučavanja

Razdvojite poznat način računanja od razumljivog objašnjenja odluke.

Pitanje 5

PROVERA ZNANJA

Koji izbor alata je najrazumniji za malu jednokratnu analizu?

A Uvek distribuirani klaster

B Excel ako pouzdano rešava zadatak

C Duboka neuronska mreža

D Najpopularniji alat na društvenim mrežama

Dobar izbor minimizuje složenost uz očuvanje pouzdanosti.

Praktični zadatak: pad prodaje

USE CASE

SITUACIJA

1. prodaja −12%
2. duže vreme dostave
3. više reklamacija
4. negativni komentari
5. više napuštenih korpi

ZADATAK

1. razdvojite kvantitativne i kvalitativne podatke
2. odredite šta može da pokaže BI analiza
3. izaberite jedan odgovarajući grafikon
4. napišite poruku za menadžment
5. predložite konkretnu aktivnost
6. procenite da li je ovo automatski Big Data

Od podataka do odluke: objasnite svaki korak, ne samo konačan odgovor.