ALATI ZA OBRADU PODATAKA

Vrste podataka i nivoi merenja

Kako pravilno klasifikovati promenljivu pre izbora grafikona i statističke metode

Predavač: dr Dušan Stefanović

TIP PODATAKA

kategorijski ili numerički

NIVO MERENJA

nominalni • ordinalni intervalni • racio

Šta treba razumeti

ISHODI UČENJA

1. razliku između kategorijskih i numeričkih podataka
2. razliku između diskretnih i kontinuiranih promenljivih
3. značenje nominalnog i ordinalnog nivoa
4. značenje intervalnog i racio nivoa
5. zašto kontekst određuje klasifikaciju promenljive
6. koji grafički prikaz odgovara kom tipu podataka

Vrsta promenljive određuje način analize

ZAŠTO KLASIFIKUJEMO PODATKE

VRSTA PREVOZA

autobus
automobil
bicikl
pešice

stubičasti dijagram

VREME PUTOVANJA

18,4 min
22,7 min
31,2 min
15,9 min

histogram ili kutijasti dijagram

Pogrešna klasifikacija vodi ka pogrešnom grafikonu, proračunu ili zaključku.

Podatke posmatramo iz dva ugla

MAPA KLASIFIKACIJE

1. PREMA TIPU PODATKA

Kategorijski

Numerički
- diskretni
- kontinuirani

2. PREMA NIVOU MERENJA

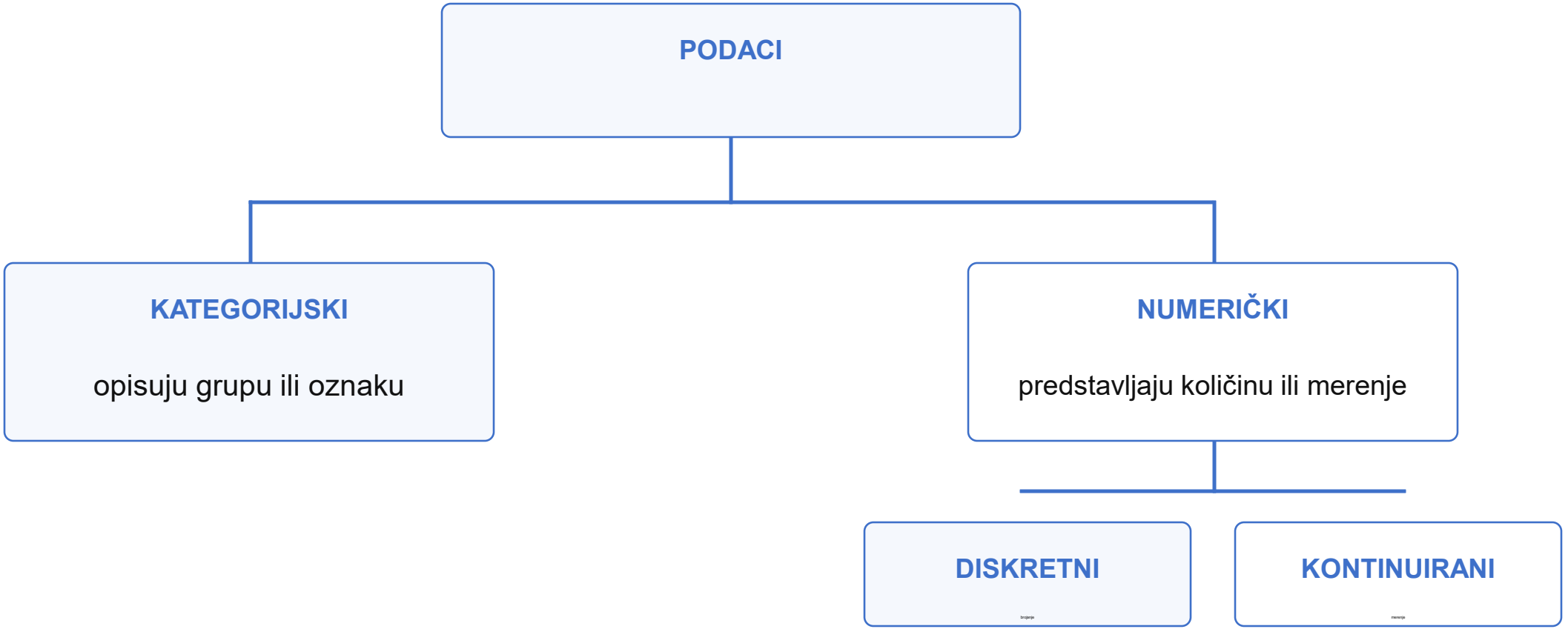
Kvalitativni
- nominalni
- ordinalni

Kvantitativni
- intervalni
- racio

Ista promenljiva dobija opis u obe klasifikacije.

Osnovna podela: kategorijski i numerički podaci

TIPOVI PODATAKA



Kategorijski podaci opisuju pripadnost grupi

KATEGORIJSKI PODACI

Vrednost odgovara na pitanje: kojoj kategoriji objekat pripada?

- AUTOBUS
- AUTOMOBIL
- BICIKL
- PEŠICE
- LIKOVNA AKTIVNOST
- MUZIČKA AKTIVNOST

Kategorije mogu biti zapisane rečima ili kodovima, ali kod nije količina.

Numerički podaci predstavljaju količinu ili merenje

NUMERIČKI PODACI

BROJANJE

12 vozila
18 dece
3 kašnjenja

najčešće diskretno

MERENJE

24,6 minuta
102,4 cm
18,7 °C

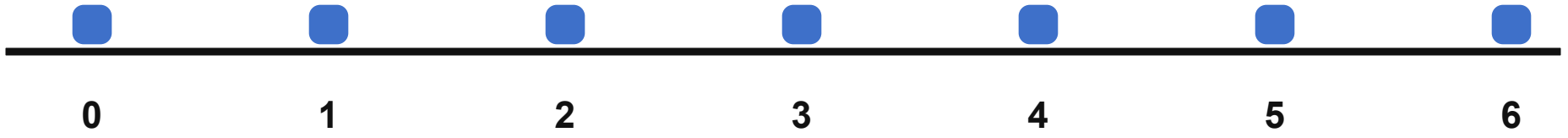
najčešće kontinuirano

Broj ima statističko značenje samo ako predstavlja merljivu količinu.

Diskretne vrednosti možemo prebrojati

DISKRETNI PODACI

Broj vozila koja su prošla raskrsnicu za jedan minut



Moguće vrednosti: 0, 1, 2, 3, 4, 5, 6 ...

Ne očekujemo 2,4 vozila u jednom posmatranom minutu.

Kontinuirana promenljiva može imati svaku vrednost u intervalu

KONTINUIRANI PODACI

Vreme putovanja od kuće do Akademije



Između 18,4 i 18,5 postoji beskonačno mnogo mogućih vrednosti.

Uređaj prikazuje zaokruženu vrednost, dok je pojava konceptualno kontinuirana.

Brojanje i merenje daju različite numeričke podatke

DISKRETNI I KONTINUIRANI

Osobina	DISKRETNI	KONTINUIRANI
Kako nastaje	brojanjem	merenjem
Moguće vrednosti	odvojene vrednosti	svaka vrednost u intervalu
Primer iz saobraćaja	broj vozila	vreme putovanja
Primer kod vaspitača	broj dece	visina deteta

Kontekst rešava granične slučajeve

ČESTA POGREŠNA TUMAČENJA

OCENE

Broj bodova 0–100 je diskretan broj.

Ocene A, B, C ili 1–5 najčešće su ordinalne kategorije.

NOVAC

U računovodstvu se beleži u najmanjoj valutnoj jedinici.

U modelima se ponekad aproksimira kao kontinuiran.

VREME

Trajanje je kontinuirano.

Vreme zabeleženo na cele minute ima diskretan zapis.

Ne klasifikujemo samo zapis. Klasifikujemo značenje i način nastanka vrednosti.

Četiri nivoa merenja određuju dozvoljena poređenja

NIVOI MERENJA

KVALITATIVNI

NOMINALNI

razlika između kategorija
bez redosleda

ORDINALNI

kategorije imaju redosled
razmaci nisu poznati

KVANTITATIVNI

INTERVALNI

jednake razlike
bez pravog nultog nivoa

RACIO

jednake razlike
i pravi nulti nivo

razlikovanje → redosled → jednake razlike → pravi nulti nivo

Viši nivo dodaje osobinu, ali izbor nivoa uvek zavisi od značenja promenljive.

Nominalne kategorije nemaju prirodan redosled

NOMINALNI NIVO

Primer: vrsta vozila u evidenciji parkirališta

AUTOMOBIL

MOTOCIKL

AUTOBUS

KOMBI

SMISLENO

broj i procenat po kategoriji

NIJE SMISLENO

prosek numeričkih kodova kategorija

Promena redosleda kategorija ne menja njihovo značenje.

Ordinalne kategorije imaju redosled

ORDINALNI NIVO

Ocena zadovoljstva organizacijom prevoza

1	2	3	4	5
veoma nezadovoljan	nezadovoljan	neutralan	zadovoljan	veoma zadovoljan

redosled postoji

razmak između susednih kategorija nije poznat

Brojevi 1–5 često predstavljaju uređene oznake, a ne precizne jednake razmake.

Intervalna skala ima jednake razlike, ali nema pravi nulti nivo

INTERVALNI NIVO

Temperatura izražena u stepenima Celzijusa



Razlika od 10 °C ima isto značenje duž skale.

0 °C ne znači odsustvo temperature.

Racio skala ima pravi nulti nivo

RACIO NIVO

Pređeno rastojanje tokom putovanja



PRAVI NULTI NIVO

0 km znači da rastojanje nije pređeno

SMISLEN ODNOS

20 km je dvostruko više od 10 km

Temperaturna skala pokazuje značaj prave nule

INTERVAL I RACIO

CELZIJUS

10 °C i 20 °C

Razlika je 10 °C.
Ne kažemo da je 20 °C dva puta toplije.

KELVIN

100 K i 200 K

Razlika je 100 K.
200 K je dvostruka apsolutna temperatura od 100 K.

Celzijus je intervalna skala. Kelvin ima pravi nulti nivo i racio osobine.

Jedan skup podataka sadrži više vrsta promenljivih

PRIMER KLASIFIKACIJE

Promenljiva	Primer vrednosti	Tip	Nivo merenja
Način dolaska	autobus	kategorijski	nominalni
Ocena gužve	velika	kategorijski	ordinalni
Broj putnika	42	numerički diskretni	racio
Vreme putovanja	27,6 min	numerički kontinuirani	racio
Temperatura	18 °C	numerički kontinuirani	intervalni

Svaku kolonu klasifikujemo posebno, čak i kada pripada istoj tabeli.

Vrsta podatka usmerava izbor grafikona

OD PODATAKA DO VIZUELIZACIJE

KATEGORIJSKI

Stubičasti dijagram

prikazuje broj ili procenat po kategoriji

NUMERIČKI

Histogram ili kutijasti dijagram

prikazuje raspodelu vrednosti

DVE NUMERIČKE

Dijagram rasipanja

prikazuje odnos dve promenljive

Grafikon treba da odgovori na pitanje bez menjanja značenja promenljive.

Tri pitanja dovoljna su za početnu klasifikaciju

SAŽETAK

1 **Da li vrednost opisuje grupu ili količinu?**

kategorijski ili numerički

2 **Ako je broj: da li brojimo ili merimo?**

diskretni ili kontinuirani

3 **Da li postoji redosled, jednak razmak i pravi nulti nivo?**

nominalni, ordinalni, intervalni ili racio

ČESTE GREŠKE

broj tretiramo kao količinu iako je šifra

ordinalnu skalu tretiramo kao precizno merenje

0 °C tumačimo kao odsustvo temperature

Pitanje 1

PROVERA ZNANJA

Broj vozila koja prođu raskrsnicu tokom pet minuta predstavlja:

A kategorijski nominalni podatak

B numerički diskretni podatak

C numerički kontinuirani podatak

D intervalni podatak

Izaberite jedan odgovor i obrazložite ga.

Pitanje 2

PROVERA ZNANJA

Skala: nezadovoljan, neutralan, zadovoljan ima koji nivo merenja?

A nominalni

B ordinalni

C intervalni

D racio

Izaberite jedan odgovor i obrazložite ga.

Pitanje 3

PROVERA ZNANJA

Temperatura izražena u stepenima Celzijusa najčešće se klasifikuje kao:

A nominalna

B ordinalna

C intervalna

D racio

Izaberite jedan odgovor i obrazložite ga.

Pitanje 4

PROVERA ZNANJA

Koja promenljiva ima pravi nulti nivo?

A temperatura u °C

B ocena zadovoljstva 1–5

C vrsta vozila

D pređeno rastojanje u kilometrima

Izaberite jedan odgovor i obrazložite ga.

Pitanje 5

PROVERA ZNANJA

Koji par je pravilno klasifikovan?

A vrsta prevoza: kontinuirana

B vreme putovanja: nominalno

C broj dece: diskretan i racio

D temperatura u °C: racio

Izaberite jedan odgovor i obrazložite ga.