ALATI ZA OBRADU PODATAKA

Od podataka do odluke

Uvod u nauku o podacima, statistiku, Python i mašinsko učenje

PODACI

011010
42%
€
A/B

ODLUKA

Predmetni profesor: dr Dušan Stefanović

Predmetni asistent: dr Milena Nikolić

POSLOVNI PROBLEM

Kako prodavnica može da prepozna kupca koji će uskoro prestati da kupuje?

PODACI	MODEL	ODLUKA
kupovine, posete, reklamacije	traži obrasce u ponašanju	ponuda, pomoć ili promena usluge

Studenti će učiti kako se problem prevodi u podatke
kako se bira metod analize
kako se rezultat pretvara u odluku

Podaci, algoritam i zaključak

OSNOVNA LOGIKA

1 PODACI

ono što merimo i prikupljamo

2 ALGORITAM

postupak koji traži obrazac

3 ZAKLJUČAK

informacija koja menja odluku

istorija kupaca

Model nelojalnog kupca

Preporuke za zadržavanje kupca

- kupci koji dugo nisu ništa kupili češće odlaze
- kupci koji ne otvaraju poruke pokazuju manje interesovanja
- veći broj reklamacija povećava verovatnoću odlaska
- redovni kupci koji reaguju na ponude uglavnom ostaju

Sadržaj kursa

PET CELINA

1 **Svet podataka**

BI, analitika, AI i poslovi

2 **Verovatnoća**

neizvesnost i mogući ishodi

3 **Statistika**

opisivanje podataka i zaključivanje

4 **Excel, Python i Power BI**

programiranje i rad sa podacima

5 **Modeli**

regresija, klasifikacija i klasteri

Alati i uloge u timu

KO ŠTA RADI

Uloga	Glavni zadatak	Osnovno pitanje	Česti alati
Data Engineer	priprema tok i kvalitet podataka	Kako da prikupimo i pripremimo pouzdane podatke?	SQL, Python, sistemi za velike podatke
BI Analyst	objašnjava poslovanje kroz izveštaje	Šta se dogodilo u poslovanju?	SQL, Excel, Power BI, Tableau
Data Analyst	istražuje podatke i proverava ideje	Zašto se to dogodilo i šta podaci pokazuju?	Python, SQL, statistika
Data Scientist	razvija modele i predviđanja	Šta bi moglo da se dogodi i kako to možemo predvideti?	Python, statistika, ML

Nazivi poslova se razlikuju, ali kvalitet podataka i razumevanje problema ostaju zajednički.

Analiza i analitika

PROŠLOST I BUDUĆNOST

ANALIZA

Šta se dogodilo

Zašto je prodaja pala u junu?

ANALITIKA

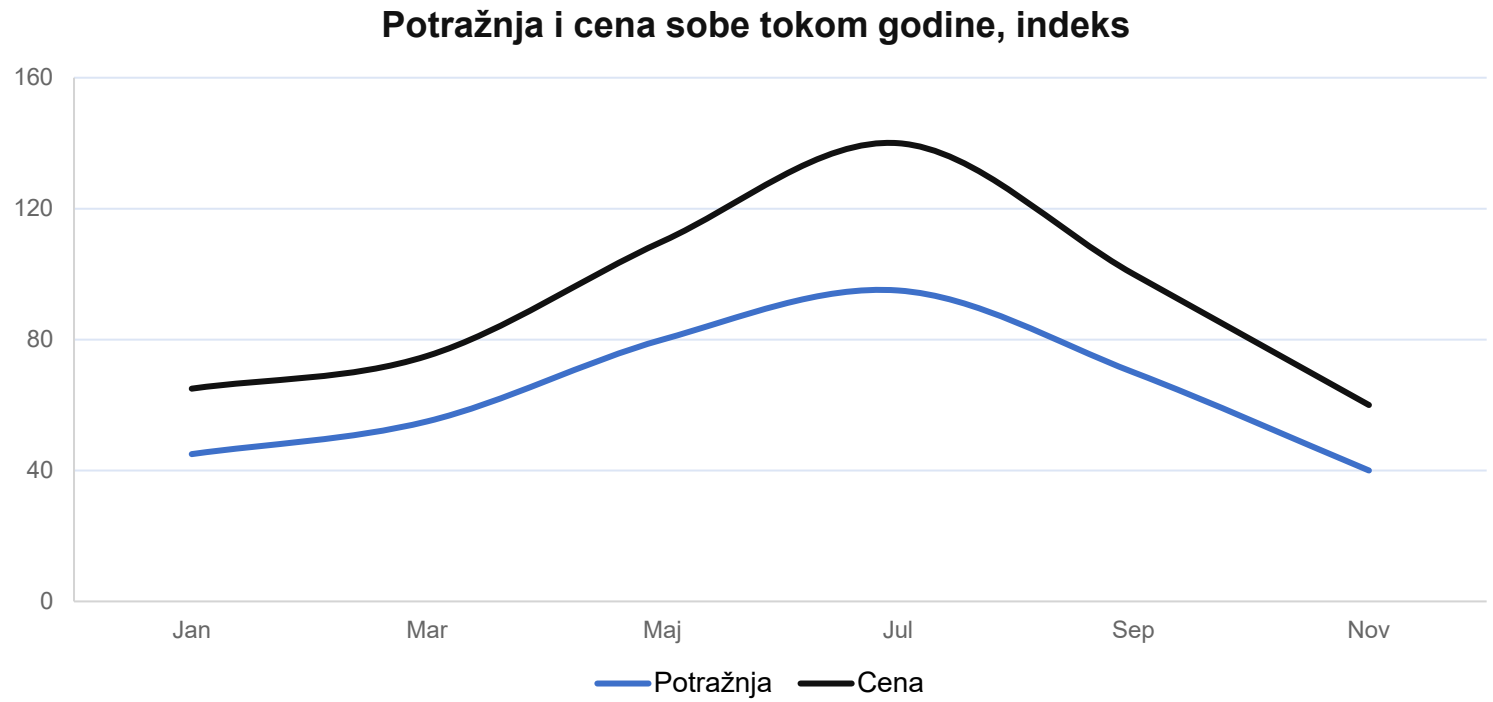
Šta bi moglo da se dogodi

Kolika će biti prodaja sledećeg juna?

Jedan pogled objašnjava prošlost, drugi podržava buduću odluku.

Primer: cena hotelske sobe

POSLOVNA INTELIGENCIJA (BUSINESS INTELLIGENCE) – VIZUALIZACIJA PODATAKA



Kada tražnja raste, hotel može povećati cenu. Kada tražnja opada, niža cena privlači goste.

Podaci pomažu da se cena prilagodi pravom trenutku.

Kada podaci postanu veliki

BIG DATA

Jedna tabela	Milioni signala	Odluka u realnom vremenu
kupci i njihove kupovine	klikovi, lokacije, senzori, video i logovi	reakcija dok se događaj još odvija

Big data ne znači samo mnogo redova. Važni su brzina, raznovrsnost i način obrade.

Tradicionalna i generativna veštačka inteligencija

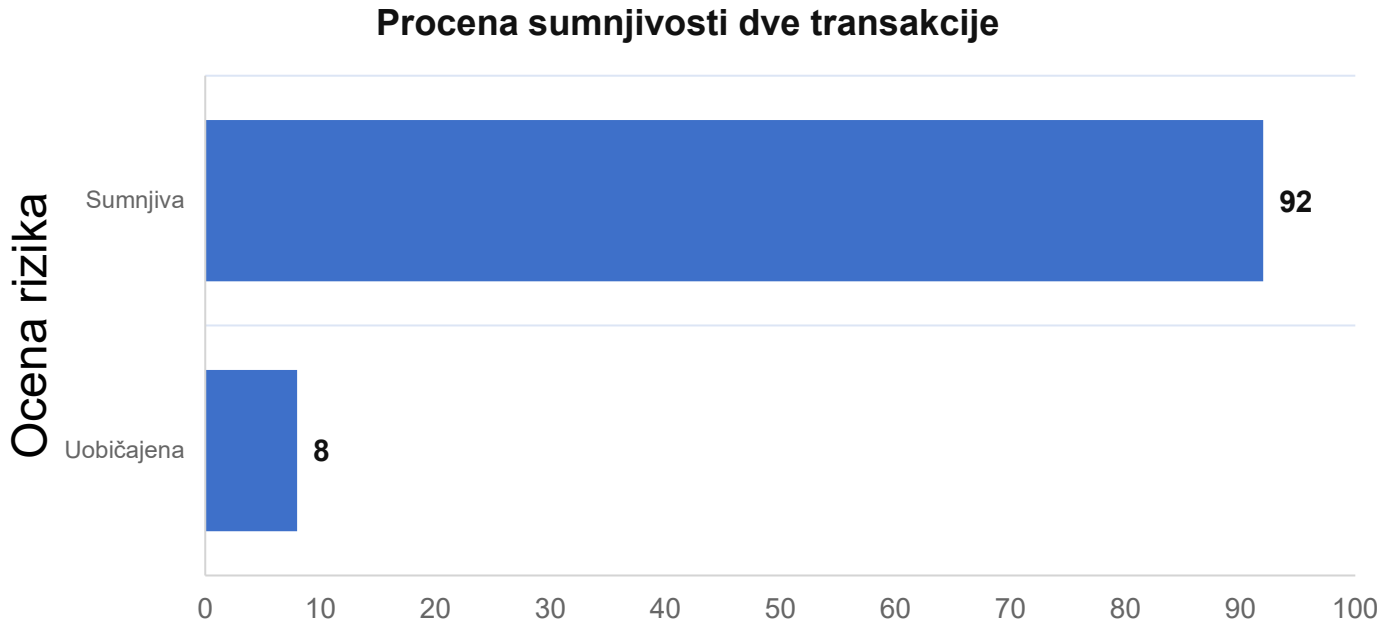
AI U PRAKSI

Pristup	Tipičan zadatak	Primer
Tradicionalni AI	prepoznaje ili predviđa	da li je transakcija sumnjiva
Generativni AI	stvara novi sadržaj	tekst, slika, kod ili sažetak

Oba pristupa koriste podatke i modele, ali rešavaju različite probleme.

Primer: otkrivanje prevare

MAŠINSKO UČENJE



Model uči iz ranijih transakcija koje su označene kao regularne ili prevarne.

Kada rizik pređe prag, banka proverava ili zaustavlja transakciju.

Verovatnoća u svakodnevnim odlukama

NEIZVESNOST

0,5

glava pri bacanju pravilnog novčića

1 : 175.000.000

primer veoma male šanse na lutriji

3%

procenjena verovatnoća kvara uređaja

Verovatnoća pretvara neizvesnost u broj koji možemo da poredimo i koristimo.

Zašto je teško dobiti na lutriji

KOMBINATORIKA

6 brojeva

iz velikog skupa mogućih brojeva

ogroman broj kombinacija

jedan dobitni izbor

među svim mogućim kombinacijama

Permutacije i kombinacije pomažu da prebrojimo ishode koje ne možemo ručno da napišemo.

Statistika pretvara tabelu u priču

DESKRIPTIVNA STATISTIKA

Kupac	Iznos kupovine	Vreme na sajtu
A	35 €	4 min
B	120 €	18 min
C	48 €	7 min
D	210 €	22 min

PROSEK

MEDIJANA

RASIPANJE

Sa nekoliko mera brzo dobijamo pregled tipičnih i neuobičajenih vrednosti.

Vizuelizacija otkriva obrazac

DVE PROMENLJIVE



Svaka tačka predstavlja jednog kupca.

Položaj tačke pokazuje odnos između dve osobine.

Kasnije ćemo slične kupce grupisati u klastere.

Od uzorka do zaključka

INFERENCIJALNA STATISTIKA

1.000 odgovora

anketa izabrane grupe kupaca

PROCENA

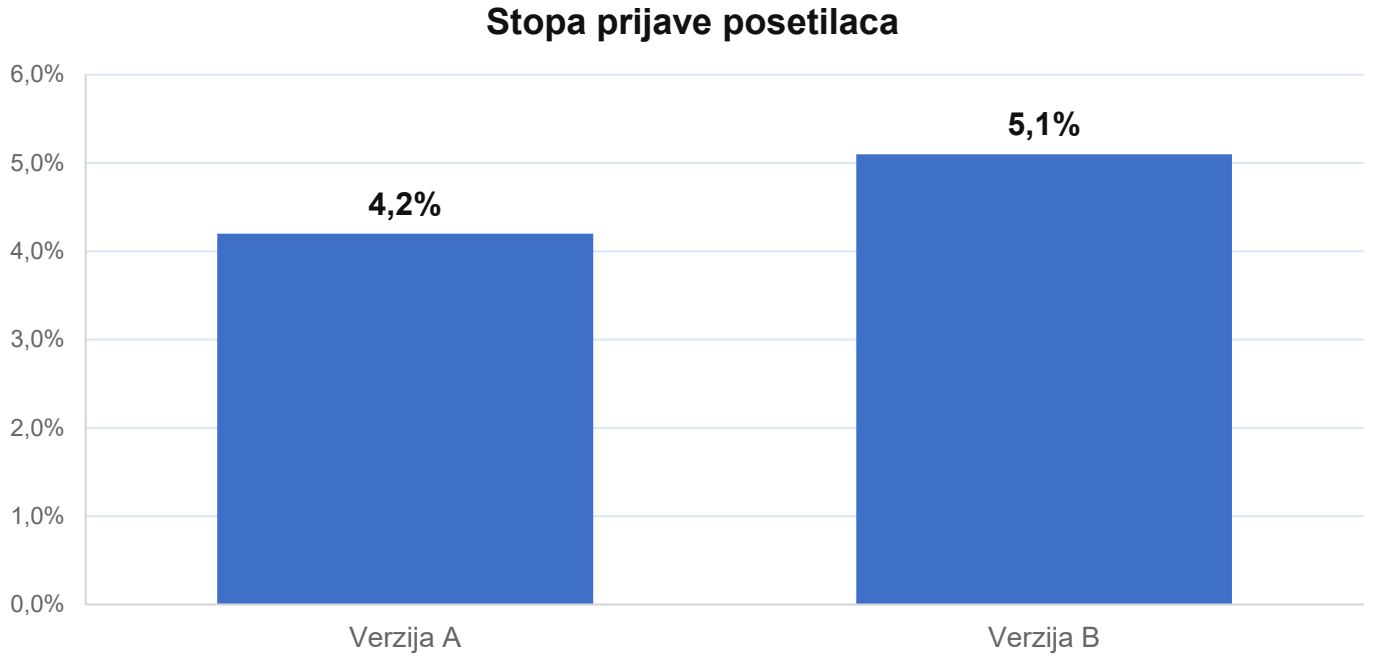
1.000.000 kupaca

šira populacija o kojoj zaključujemo

Uzorak štedi vreme, ali zaključak uvek nosi određenu neizvesnost.

Primer: koji dizajn sajta je bolji

A/B TESTIRANJE



Grupa A vidi prvi dizajn. Grupa B vidi drugi dizajn.

Statistika pomaže da proverimo da li je razlika stvarna ili slučajna.

Python i PowerBI

PROGRAMIRANJE

```
import pandas as pd

podaci = pd.read_csv('prodaja.csv')
print(podaci.head())
print(podaci['iznos'].mean())
```

Učitavanje podataka

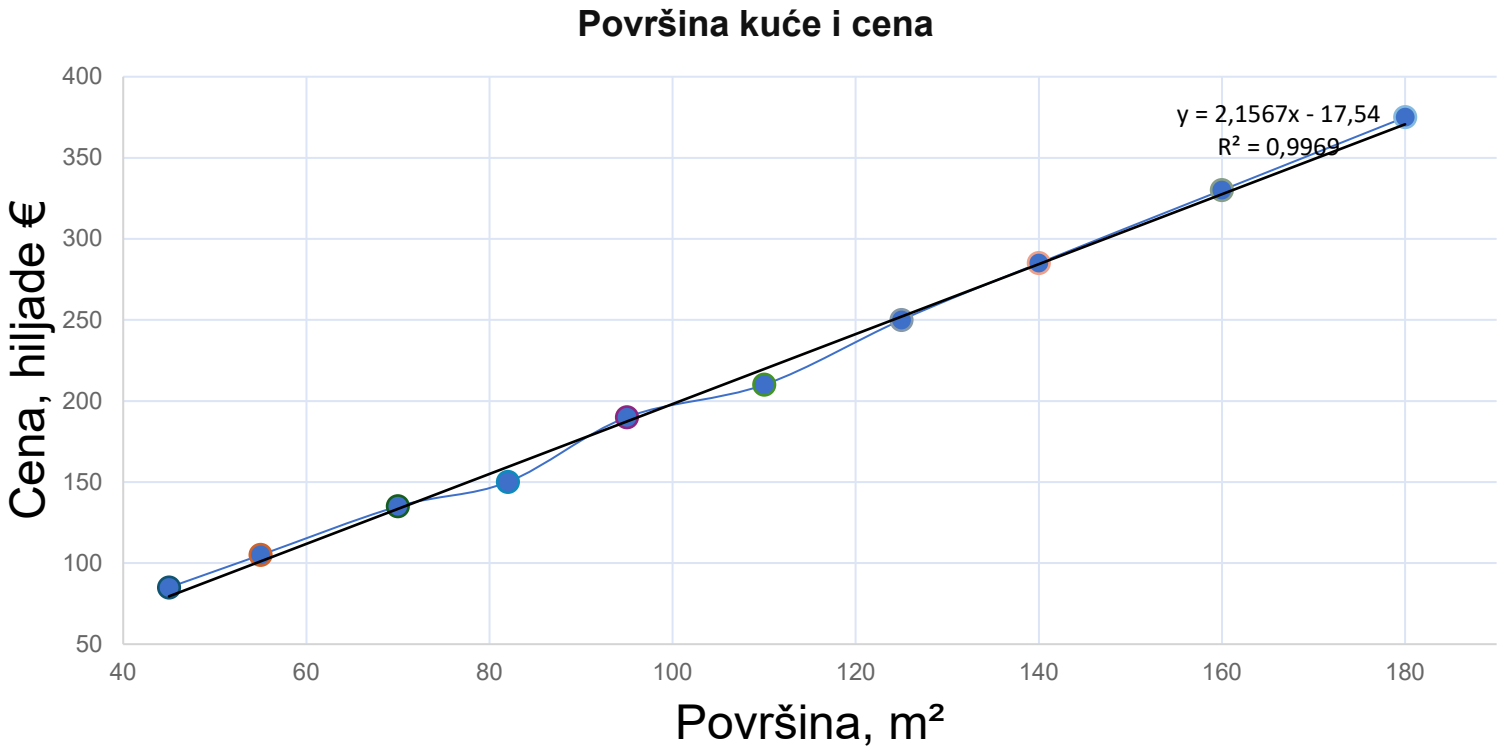
Pregled tabele

Prvi proračun

Jupyter povezuje kod, rezultate, grafikone i objašnjenje u jednom dokumentu.

Primer: procena cene kuće

LINEARNA REGRESIJA



Veće kuće uglavnom imaju višu cenu.

Model uči vezu iz poznatih primera i procenjuje cenu nove kuće.

Primer: da li će kupac otići

LOGISTIČKA REGRESIJA

PODACI O KUPCU

učestalost kupovine
reklamacije
trajanje članstva
reakcija na ponude

MODEL

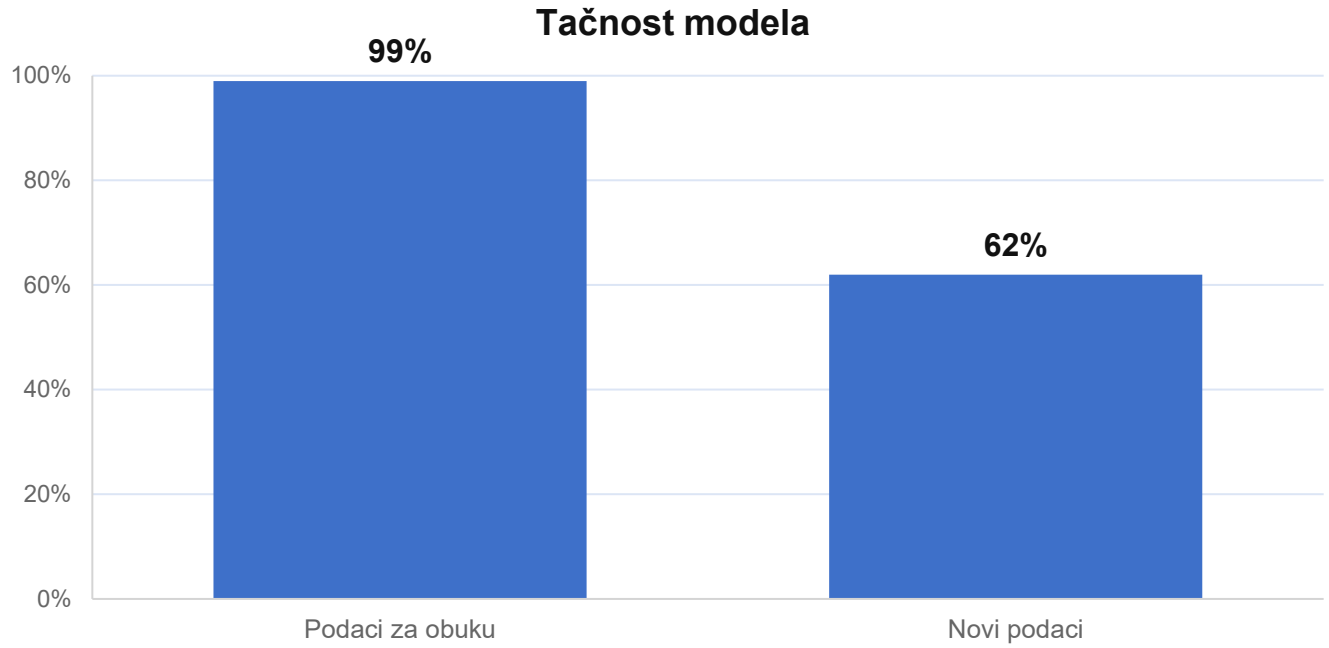
78%

procenjena verovatnoća odlaska

Rezultat podržava odluku kome treba ponuditi pomoć ili pogodnost.

Kada model nauči primere napamet

OVERFITTING

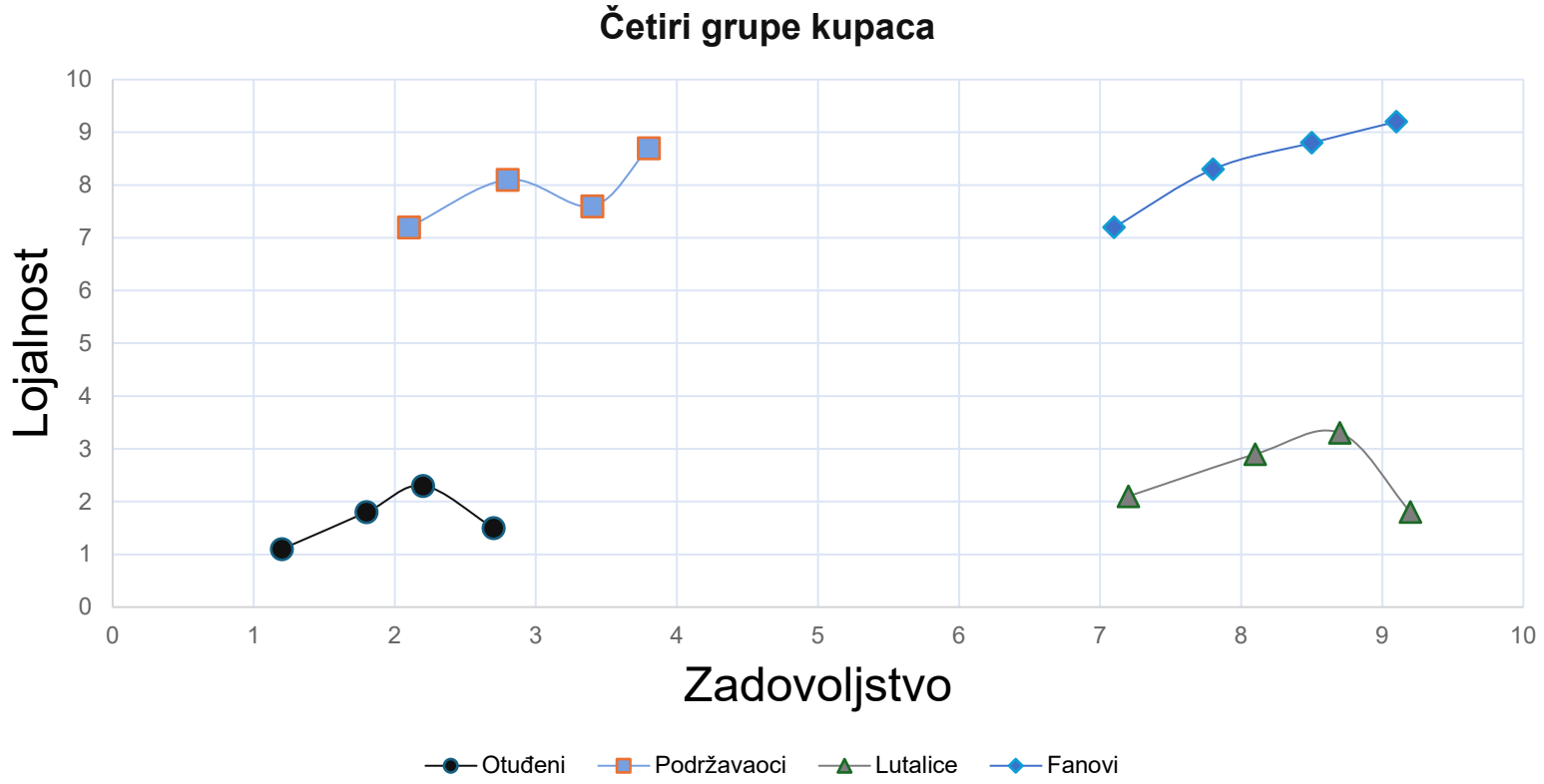


Model izgleda odličan dok gleda poznate primere.

Pravi test počinje kada dobije podatke koje ranije nije video.

Kupci se prirodno grupišu

KLASTER ANALIZA



Klasterisanje ne traži unapred zadate oznake.

Algoritam pronalazi kupce sa sličnim osobinama.

Svaka grupa može dobiti drugačiju ponudu.

Šta ćete moći da uradite

ISHODI KURSA

1 **POSTAVITI DOBRO PITANJE**

od poslovnog problema do analize

2 **PRIPREMITI I PRIKAZATI PODATKE**

tabele, grafikoni i opisne mere

3 **PROCENITI NEIZVESNOST**

verovatnoća i statističko zaključivanje

4 **PISATI PYTHON KOD**

Jupyter, funkcije, petlje i paketi

5 **IZGRADITI I PROVERITI MODEL**

regresija, klasifikacija i klasteri

Najvažniji rezultat je sposobnost da iz podataka izvedete razumljivu i korisnu odluku.

Putovanje počinje

ŠTA NAS ČEKA

OTKRITI PREVARU	UPOREDITI DVA DIZAJNA	PREDVIDETI CENU	GRUPISATI KUPCE
transakcije	A/B eksperiment	regresija	klaster analiza