

The background features a series of horizontal blue lines on the left that transition into a complex, tangled web of orange lines on the right, creating a visual metaphor for the transition from structured rules to complex, interconnected patterns.

Od pravila do obrazaca: Kako podaci menjaju programiranje

Uvod u mašinsko učenje i programiranje na osnovu podataka.

Kako biste rešili ovaj problem?

Pokušajte da zaključite iz brojeva koje imate u levoj koloni (**ulazi**) kako se oni odnose na brojeve u desnoj koloni (izlazi).

ULAZ		IZLAZ	
1, 9, 5, 4			19
3, 8, 11			22
6, 7, 9, 2, 2			26
2, 3, 6			11

Rešenje je jasan skup pravila.

Verovatno ste brzo uočili da izlaz predstavlja zbir brojeva na ulazu. Za ovaj zadatak, lako je napisati algoritam koji vas vodi do rešenja.



```
numbers = [1, 9, 5, 4]
sum = 0
for element in numbers:
    sum = sum + element
```

A kako biste rešili ovaj?

Sada se okušajte u sledećoj mozgalici. Ulazi su sada fotografije životinja, a izlazi su brojevi 0 ili 1.

ULAZ	IZLAZ
	1
	0
	0

Zid složenosti

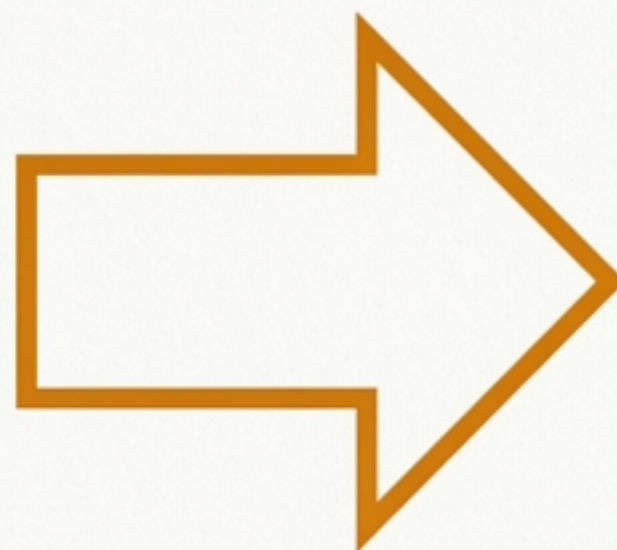
Rešenje je prepoznati da li je na fotografiji mačka (izlaz 1) ili ne (izlaz 0). Ali kako biste za to napisali algoritam? Ne zaboravite da uzmete u obzir pozadinu, osvetljenje, ugao gledanja i činjenicu da postoji preko 50 različitih vrsta mačaka.

Postoje problemi gde, čak i ako se potrudimo da napišemo hiljade pravila, nismo u stanju da kreiramo precizne algoritme.

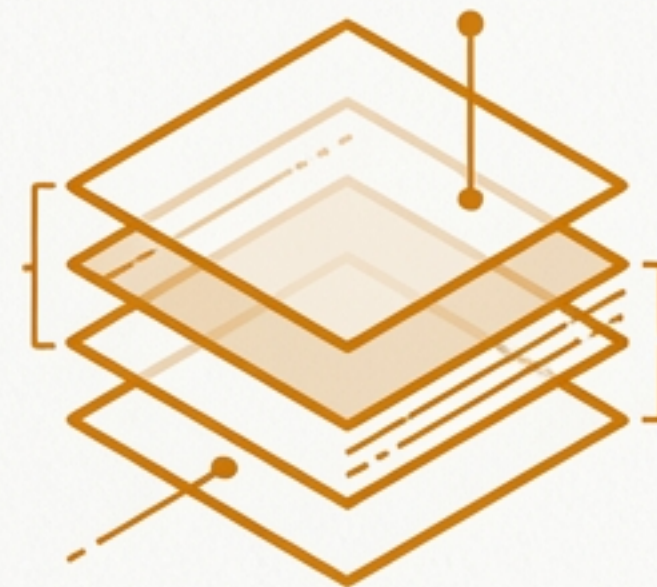


Ako ne možemo napisati pravila, učimo iz primera.

Eksplicitna pravila

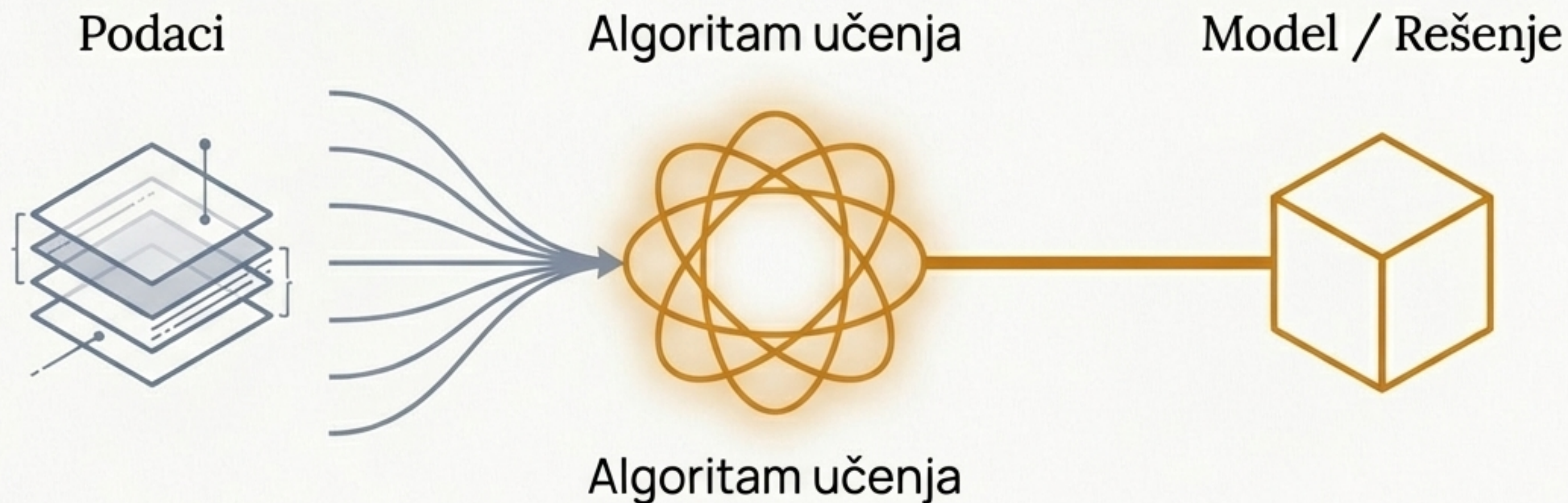


Učenje iz primera



Probleme poput ovog ne rešavamo pisanjem eksplicitnih pravila, već ih opisujemo pomoću skupova podataka – parova ulaza i očekivanih izlaza. To je suština programiranja na osnovu podataka.

Dobrodošli u programiranje na osnovu podataka



To je pristup koji naglašava upotrebu podataka kao primarnog pokretača za razvoj algoritama. Umesto da se oslanjamo isključivo na unapred definisanu logiku, programeri se fokusiraju na prikupljanje, analizu i korišćenje podataka kako bi informisali rešenje.

Podaci: "Gorivo" koje pokreće moderne algoritme

Efikasnost i tačnost sistema mašinskog učenja u velikoj meri zavise od dostupnosti velikih količina visokokvalitetnih podataka. Veliki skupovi podataka omogućavaju:



Poboljšana tačnost

Modeli obučeni na većim skupovima podataka imaju tendenciju da bolje generalizuju.



Robusnost

Raznovrsni podaci omogućavaju modelima da efikasnije rukuju rubnim slučajevima i anomalijama.



Adaptivnost

Sistemi se prilagođavaju promenljivim uslovima kontinuiranim učenjem iz novih podataka.



Automatizacija

Mašine uče iz istorijskih podataka i predviđaju bez ljudske intervencije.

Anatomija skupa podataka

Skup podataka je **strukturirana zbirka podataka** koja se koristi za **obuku i evaluaciju modela mašinskog učenja**.

Skup podataka (Dataset)	Kvadratura (m ²)	Lokacija (Zona)	Broj soba	Cena (€)	Atribut (Attribute): Pojedinačna varijabla ili karakteristika koja opisuje instancu. Instanca (Instance): Jedan zapis ili posmatranje unutar skupa podataka.
	75	1	3	150,000	
	120	1	5	280,000	
	50	2	2	95,000	
	90	3	4	110,000	

Šta mašina gleda, a šta pogađa?

Atributi se dele na dve ključne uloge:

Kvadratura (m ²)	Lokacija (Zona)	Broj soba	Cena (€)	Cena (€)
75	1	3		150,000
120	1	5		280,000
50	2	2		95,000
90	3	4		110,000



Ulazne varijable (Input Variables)

Specifični atributi koji se koriste kao prediktori. Pružaju neophodne informacije modelu da nauči obrasce.

****Primer**:** U modelu predviđanja cena stanova, ulazne varijable mogu biti `kvadratura` i `lokacija`.

Izlazne varijable (Output Variables)

Ciljne vrednosti koje model ima za cilj da predvidi na osnovu ulaznih varijabli.

****Primer**:** U istom modelu, izlazna varijabla bi bila `procenjena cena`.

Šta je 'Model' u mašinskom učenju?

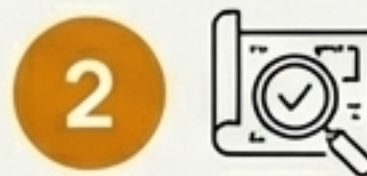


Model je matematička reprezentacija koja mapira date ulazne varijable na izlazne varijable na osnovu naučenih obrazaca iz podataka o obuci.



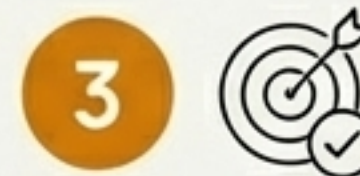
1. Obuka (Training)

Model uči iz podataka podešavanjem svojih parametara kako bi se smanjile greške predviđanja.



2. Mapiranje (Mapping)

Jednom obučen, model može generisati predviđanja za nove ulaze.

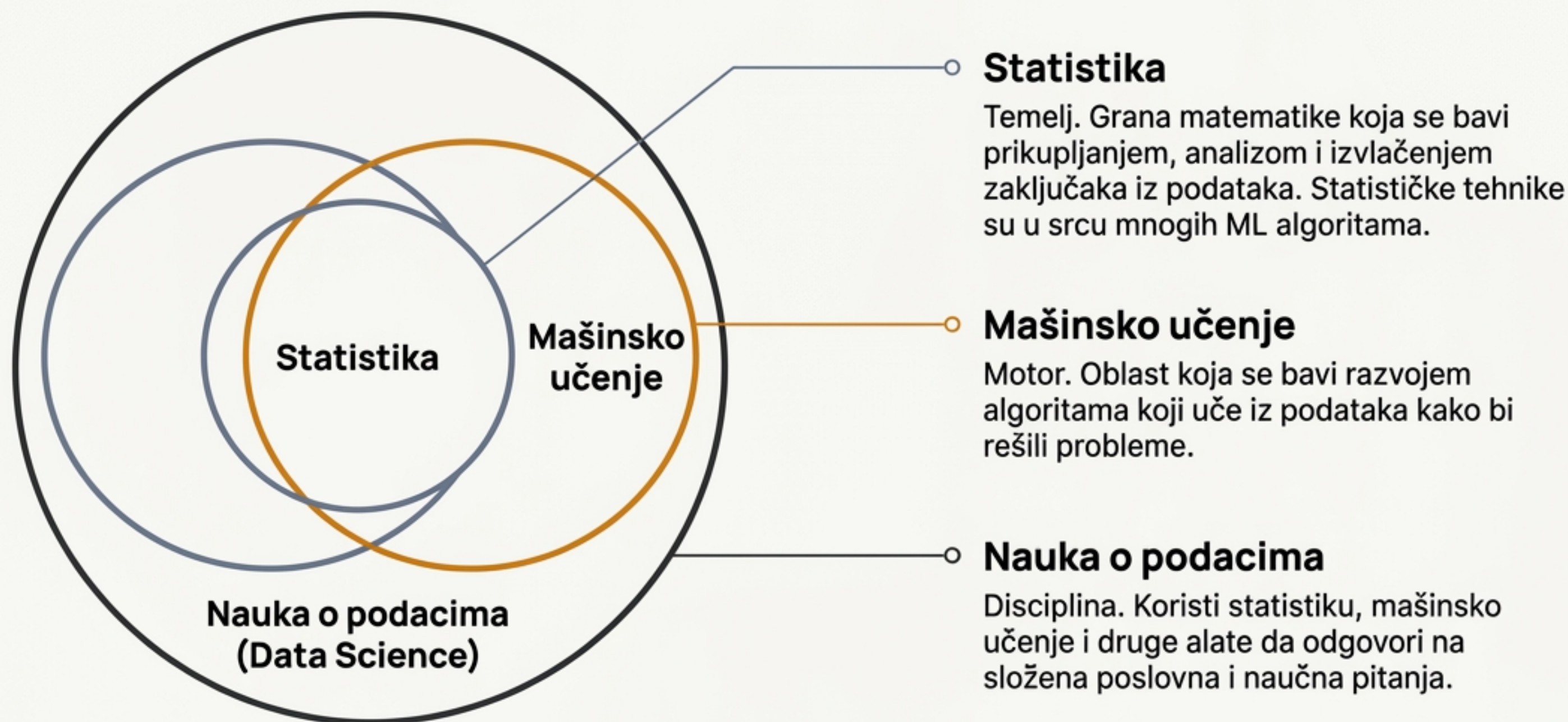


3. Evaluacija (Evaluation)

Performanse modela se procenjuju korišćenjem metrika kao što su tačnost, preciznost, i druge.

Gde se mašinsko učenje uklapa?

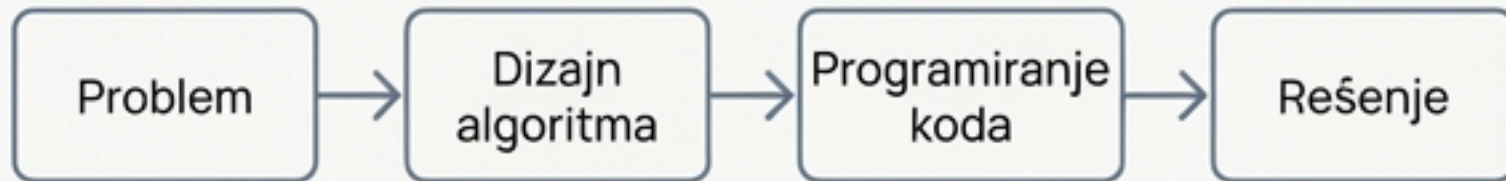
Mašinsko učenje je moćan alat, ali je deo šireg polja koje se bavi podacima.



Dva puta do rešenja: Klasično vs. Na osnovu podataka



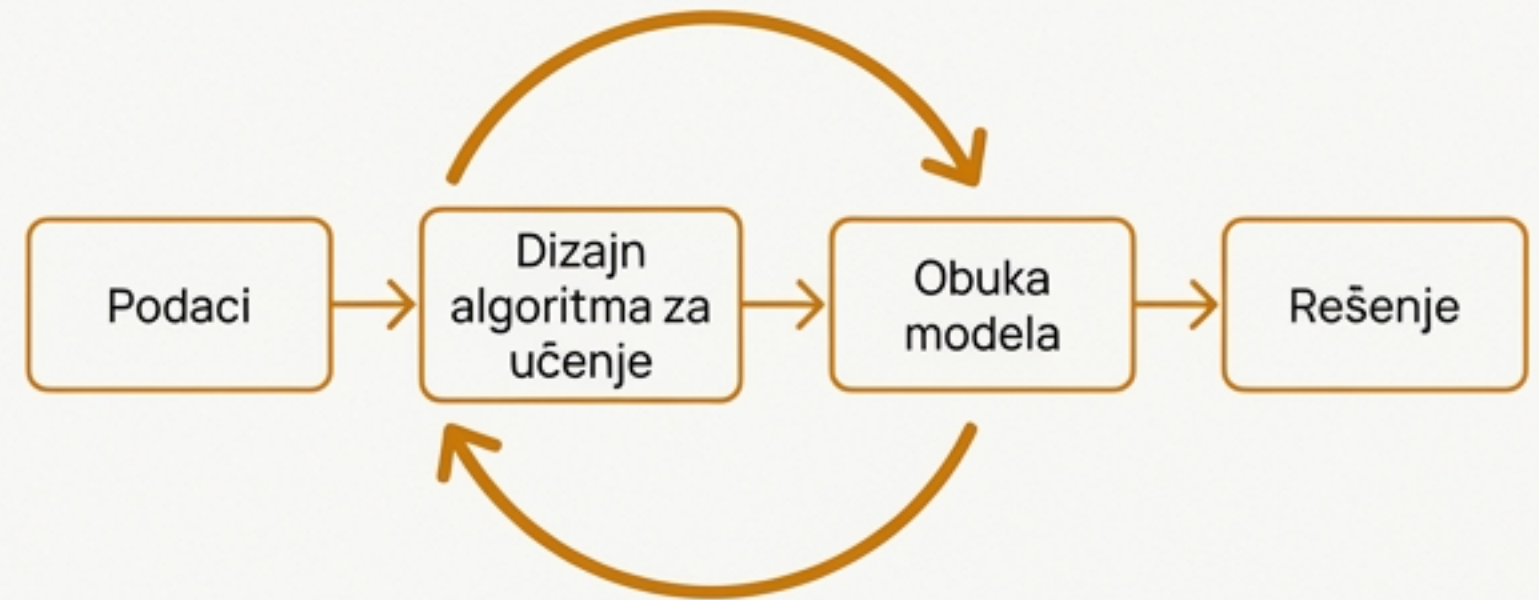
Klasično programiranje



Dizajniramo algoritam za rešavanje problema.
Rešenje je statično.



Programiranje na osnovu podataka



Dizajniramo algoritam koji *uči kako* da reši problem iz podataka. Ovaj proces je inherentno iterativan.

Ključne ideje za poneti



1

Neki problemi su previše složeni za eksplicitna pravila.

Problemi poput prepoznavanja slika ne mogu se efikasno rešiti ručnim pisanjem pravila.



2

Programiranje na osnovu podataka omogućava sistemima da uče obrasce direktno iz primera.

Ovo je rešenje za složene probleme, gde logika proizilazi iz podataka.



3

Model je rezultat tog učenja – mapa koja povezuje ulaze sa izlazima.

Centralni artefakt nije kod sa pravilima, već matematička reprezentacija naučenih obrazaca.



Hvala.

Pitanja i diskusija