

Predavač: Jelena Slivka slivkaje@uns.ac.rs

Asistenti: Dragan Vidaković vdragan@uns.ac.rs,

Katarina-Glorija Grujić katarina.glorija@uns.ac.rs

Mašinsko učenje

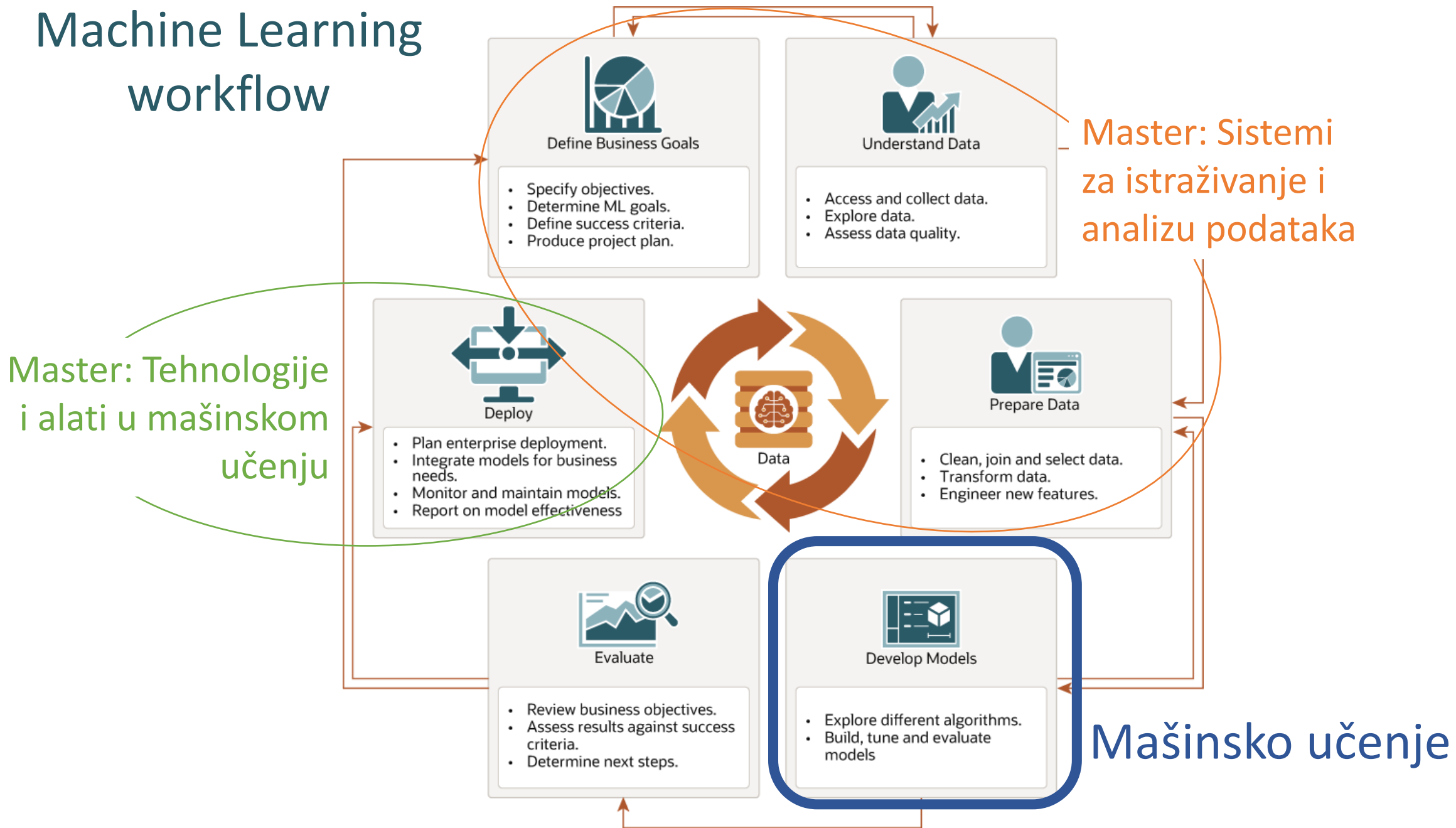


Šta radimo?

Kako radimo?

Zašto je to bitno?

Machine Learning workflow



Šta radimo?



Develop Models

- Explore different algorithms.
- Build, tune and evaluate models

**Fokus:
Tradicionalni
modeli**

Identifikacija tipa problema

Matematičke osnove ML modela – razumevanja njihovih prednosti i ograničenja

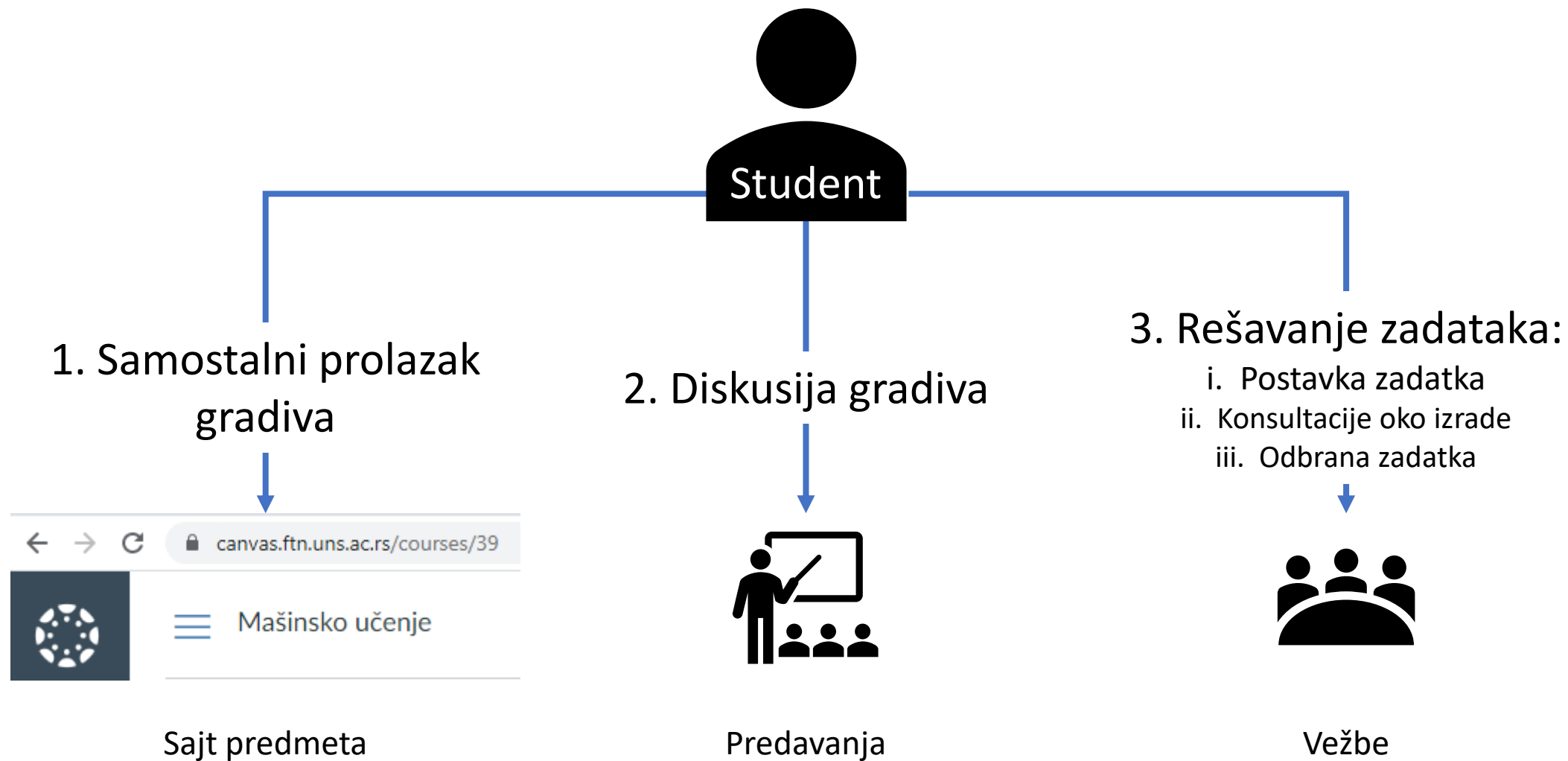
Selekcija adekvatnog modela za dati problem

Evaluacija modela

Praktični saveti za primenu mašinskog učenja

Teorija učenja

Kako radimo?



Samostalni prolazak gradiva

<https://canvas.ftn.uns.ac.rs/courses/39>

← → ↻ canvas.ftn.uns.ac.rs/courses/39

Mašinsko učenje

Home

Announcements

Assignments

Discussions

Grades

People

Pages

Files

Syllabus

Outcomes

Rubrics

Quizzes

Modules

Mašinsko učenje

[Ovde](#) možete pronaći informacije vezane za polaganje predmeta.

R.br.	Datum (sreda)	Predavanja
1	22.02.2023.	Uvodno predavanje
2	01.03.2023.	Predavanje 1: Jednostruka linearna regresija
3	08.03.2023.	Predavanje 2: Višestruka linearna regresija
4	15.03.2023.	Predavanje 3: Regularizacija
5	22.03.2023.	Predavanje 4: Regularizacija (nastavak) i Neparаметarski pristup
6	29.03.2023.	Predavanje 5: Metod maksimalne verodostojnosti (Maximum Likelihood)
		Predavanje 6: Linearni modeli za klasifikaciju

Announcements

Assignments

Discussions

Grades

People

Pages

Files

Syllabus

Outcomes

Rubrics

Quizzes

Modules

Settings

Predavanje 1: Jednostruka linearna regresija

1. Uvod
https://www.youtube.com/watch?v=IbNmd8IJ3U&ab_channel=slivkaje
2. Model i interpretacija
<https://www.youtube.com/watch?v=8M2FhuD1Nl4&feature=youtu.be>
3. Obučavanje modela: Gradijentni spust
https://www.youtube.com/watch?v=R_L9O7c_28A&feature=youtu.be

Home

Announcements

Assignments

Discussions

Grades

People

Pages

Files

Syllabus

Mašinsko učenje > Files > Predavanja > Predavanje 1: Jednostruka linearna regresija

Search for files

0 items selected

Mašinsko učenje

Predavanja

- Predavanje 10: Klasterovanje
- Predavanje 11: Redukcija dimenzionalnosti
- Predavanje 12: Praktični saveti za primenu
- Predavanje 13: Teorijske osnove nadgledanja
- Predavanje 1: Jednostruka linearna regresija
- Predavanje 2: Višestruka linearna regresija

Name

- 1-1 Uvod.pdf
- 1-2 Model i njegova interpretacija.pdf
- 1-3 Obučavanje modela-Gradijentni spust.pdf

Polaganje predmeta

Teorija

- Maksimum 40 bodova
- Minimum 15 bodova da bi se teorija položila (bodovi iz teorije računali)
- Predmet možete položiti i isključivo putem praktičnog dela
- Polaganje: test u FTN rokovima
- Ukoliko popravljate teoriju, prethodni bodovi iz teorije vam se poništavaju

Polaganje
predmeta

Praktični deo

- Maksimum 60 bodova
- 5 domaćih zadataka i projekat
- Timski rad (2-3 studenta)
 - Svaki član mora da da svoj doprinos
 - Bodovi dodeljeni članovima istog tima mogu da se razlikuju

Polaganje predmeta

Praktični deo

Opcija 1 (maksimum 25 bodova)

- Izrada projekta 25 bodova

Opcija 2 (maksimum 50 bodova):

- Uspešno odrađena 2 domaća zadatka: 25 bodova
- Izrada projekta 25 bodova

Opcija 3 (maksimum 60 bodova):

- Uspešno odrađenih 4 ili više zadatka sa vežbi: 35 bodova
- Izrada projekta 25 bodova

Opcija 4 (maksimum 60 bodova):

- Nagrada za najuspešnije timove (koji odlično urade 4+ domaća zadatka)

Polaganje
predmeta

Praktični deo

01.03. Prijava timova

Domaći zadaci:

- 04.03. - 19.03. Jednostruka linearna regresija
- 25.03. - 16.04. Višestruka regresija
- 22.04. - 08.05. Klasifikacija
- 13.05. - 22.05. Modeli ansambla
- 27.05. – 02.06. Klasterovanje

Prijava projekta: 03.06. - 17.06.

Odbrana projekta:

- Početak jula
- Dodatni rok: početak septembra – u ovom roku se poeni osvojeni na projektu množe sa 0.7

Polaganje predmeta

Praktični deo

Domaći zadaci:

- Testiranje modela na skrivenom test skupu podataka na *online* platformi;
- Prpratni izveštaj na srpskom ili engleskom jeziku
- Zadatak je uspešno urađen ukoliko model ispuni zadati kriterijum
- Najbolji tim iz grupe prezentuju svoje rešenje na početku vežbi
- Svi koji su uradili zadatak u terminu ga brane
- **Plagijat = diskvalifikacija!!!**

Projekat:

- Sami predlažete temu
- Implementacija rešenja
- Prpratni izveštaj po definisanom šablonu

Polaganje predmeta

Praktični deo

Ostvareni rezultati i kako se do njih došlo:

- Pristup problemima
- Korišćeni algoritmi
- Određivanje hiper-parametara algoritama
- Rad sa trening skupom podataka

Propratni izveštaji:

- Sadržaj propratnih izveštaja
- Usklađenost izveštaja i izvornih kodova rešenja

Diskusija:

- Prezentovanje rešenja i odgovori na pitanja prilikom prezentovanja

Anketa 2023

Koji aspekti kursa su vam bili najmanje korisni ili frustrirajući?

Konstantno postojanje nekih rokova, kojih nije malo 🙄👉 i sam momenat potvrde rešenja na platformi jer imamo samo dva pokušaja

Na početku kursa me je frustrirala platforma za domaće zadatke jer su nam dozvoljena samo 2 *submit*-a, ali iz ove perspektive se slažem da je to manje-više dovoljno.

Vežbe, generalno. Malo je nejasno kako rangiranje na tabelama utiče na krajnje bodove i rezultat pa postoji anksioznost koja traje čitav semestar. Takođe vežbe nisu dovoljno individualne, bilo bi dobro da postoji nekakav feedback za pojedinačna rešenja, kako ih unaprediti, šta su greške, i slično.

Ako bih baš morala nešto da odaberem, možda bih odabrala to što se čini da u nekim slučajevima testni skup nije bio dovoljno reprezentativan, pa su najbolje skorove postigli timovi koji nisu uradili ništa posebno, samo su imali sreće sa parametrima. Nisam sigurna da je to zaista bilo tako, ali tako se pričalo.

Malo fleksibilniji rokovi

Broj pokušaja na platformi je uvećan na 3

Feedback za pojedinačna rešenja na vežbama

Transparentnije bodovanje (*rubrics*)

Anketa 2023

Koji aspekti kursa su vam bili najmanje korisni ili frustrirajući?

Jedan čas upoznavanja sa scikit bibliotekom bi možda bi značajan nekim studentima.

Da smo imali vezbe duze od 5 minuta, mozda bih imao nesto da dodam.

Sadržajnije vežbe

Nove verzije svih materijala, od interpretera do biblioteka. Ne mora sve nužno biti najnovije, ali dovoljno nezastarelo da može da se ubaci u PyCharm :)

Novije verzije *Python*-a i biblioteka

Anketa 2023

Koji aspekti kursa su vam bili najmanje korisni ili frustrirajući?

Termini predavanja su zanimljivi i interaktivni, ali smatram da potreba za angažovanjem maltene duplo više vremena odgovara veliki broj studenata od aktivnog učestvovanja. Pod ovim podrazumevam vreme potrošeno na klipove, koji su često duži od termina predavanja, plus samo predavanje. Iako ono ponekad ne traje ceo termin, i dalje iziskuje dolazak na fakultet, što je dodatni utrošak vremena. Ovo je izvodivo na početku semestra, ali posle nekoliko nedelja za većinu studenata jednostavno nije održivo.

Pojasniti benefite obrnute učionice

Preobimna predavanja bez izdvajanja suštine, potrebno više programerskih manje teorijskih primera

Jako mi se svidelo kada je među prvim terminima predavanja bio pokretan kod na nekom pravom problemu, lepo je videti direktnu primenu teorije. Možda bi bilo dobro više takvih primera dodati.

Veći broj studija slučaja u terminima predavanja

Anketa 2023

Koji aspekti kursa su vam bili najmanje korisni ili frustrirajući?

Smatram da samo rangiranje utice jako lose na studente.

Šta nismo adresirali

Sažimanje klipova, idealno ne samo sečenjem već kompletnom preradom, iako znam da to nije realistično zbog velike količine truda i vremena koje bi se morale uložiti u takav poduhvat

Staviti ovaj predmet pre soft kompjutinga

Svi posto me ne zanima ML :D

Malo me je frustriralo kad radimo neke matematičke dokaze koji zahtijevaju gradivo sa prve godine, iz analize, a ja sam sve zaboravila. To nisam mogla da ispratim, ali nije greška u predmetu, nego u meni



Hvala 😊

Zelim samo da pohvalim asistenta Dragana jer je bio transparentan, azuran i apsolutno uvek se trudio da pomogne i objasni studentima sve sto je trebalo ❤️

Samo zelim da pohvalim ceo tim, nisam uspela da polozim ove godine ali zaista nemam nikakve zamerke, smatram da je nacin rada bio veoma dobar i koristan. Hvala Vam na tome

Sve pohvale za celokupan kurs i mogućnost oslobađanja od usmenog dela je ogromna motivacija za studente da dobro urade domaće zadatke. Svakako, smatram da se predavanja moraju pogledati kako bi se odlično uradili zadaci, tako da je to adekvatna nagrada. Sadržaj predmeta je koristan i zanimljiv, nije opterećujuć previše i jasno se primenjuju naučeni koncepti sa predavanja bez suvišnog ponavljanja već urađenog kao što to zna da bude na mnogim kursevima. Srdačan pozdrav za nastavno osoblje!

Zašto je to bitno?

Oskudica talenta



The New York Times

2017: Tech Giants Are Paying Huge Salaries for Scarce A.I. Talent

“Demand for data scientists is off the charts”.

US-based business need more than 150,000 data scientist jobs filled

<https://economicgraph.linkedin.com/resources/linkedin-workforce-report-august-2018>

PTC

Усвојена Стратегија развоја
вештачке интелигенције

Влада Србије усвојила је Стратегију развоја вештачке интелигенције за период од 2020. до 2025. То је део развоја свеукупне Стратегије развоја науке у Србији до 2030. Србија је прва земља у Југоисточној Европи и 26. у свету која има такву стратегију. То нам обезбеђује већи економски раст и квалитетнији живот грађана,

Just How Shallow is the Artificial Intelligence Talent Pool?

Research from Element AI indicates only 22,000 have right skills globally

By [Jeremy Kahn](#)

February 7, 2018, 2:01 PM GMT+1

Google's self-training AI turns coders into machine-learning masters

TOM SIMONITE BUSINESS 10.13.17 07:00 AM

**GOOGLE'S LEARNING
SOFTWARE LEARNS TO WRITE
LEARNING SOFTWARE**

Google's AutoML lets you train custom machine learning models without having to code

Google has started using AI to build more advanced AI

Introducing Cloud AutoML

A technology that can automatically create a Machine Learning Model

Computational power could replace the need for machine learning expertise.

Google's Head of AI, Jeff Dean



We hope AutoML will make it possible in **three to five years** for hundreds of thousands of developers **to design new neural nets** for their particular needs

Google CEO, Sundar Pichai

Structured data

Language

Sight

Da li вреди učiti mašinsko učenje?

Inovacija nastaje tako što radimo stvari drugačije, a ne više



Now that you are here, do we still need ML experts?



As a language model, I can assist in certain tasks related to machine learning, but I am not a substitute for human expertise. Machine learning is a complex field that requires not only technical skills but also creativity, critical thinking, and the ability to make decisions and solve problems. Human ML experts are essential for designing and implementing innovative solutions, interpreting and explaining model behavior, and ensuring ethical and responsible use of AI systems.

Koje poslove
ML može da
unapredi?

Šta ML ne
može?

Šta nam je
strategija i cilj?

Kako da
postavimo
problem?

Šta ML
eksperti
zapravo rade?



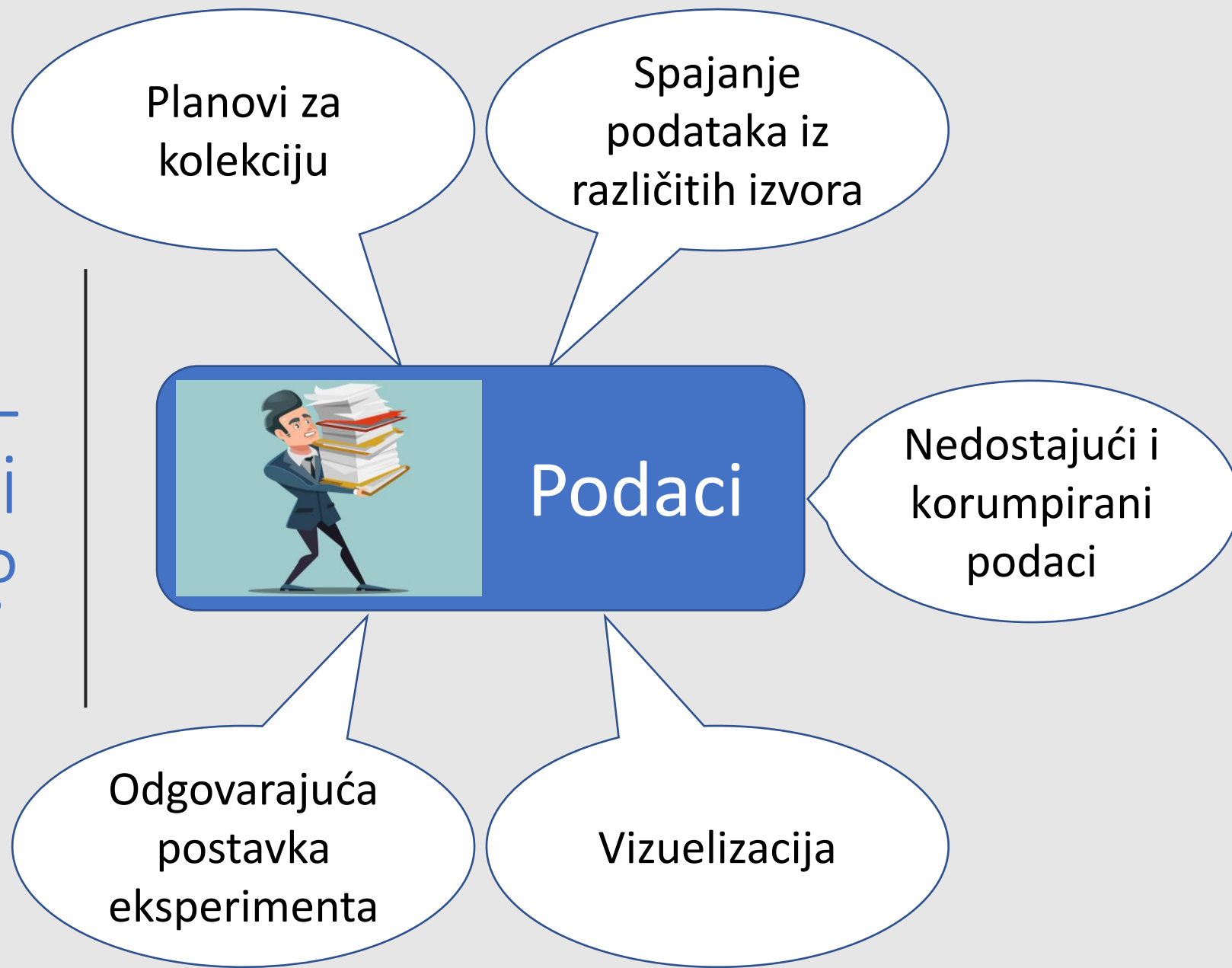
Razumeju kontekst

Šta su
operaciona
ograničenja?

Postoje li
etički rizici?

Postoje li
biasi?

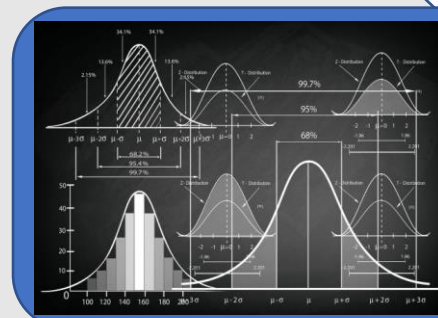
Šta ML
eksperti
zapravo rade?



Šta ML
eksperti
zapravo rade?

Selekcija
modela

Prilagođavanje
modela
ograničenjima

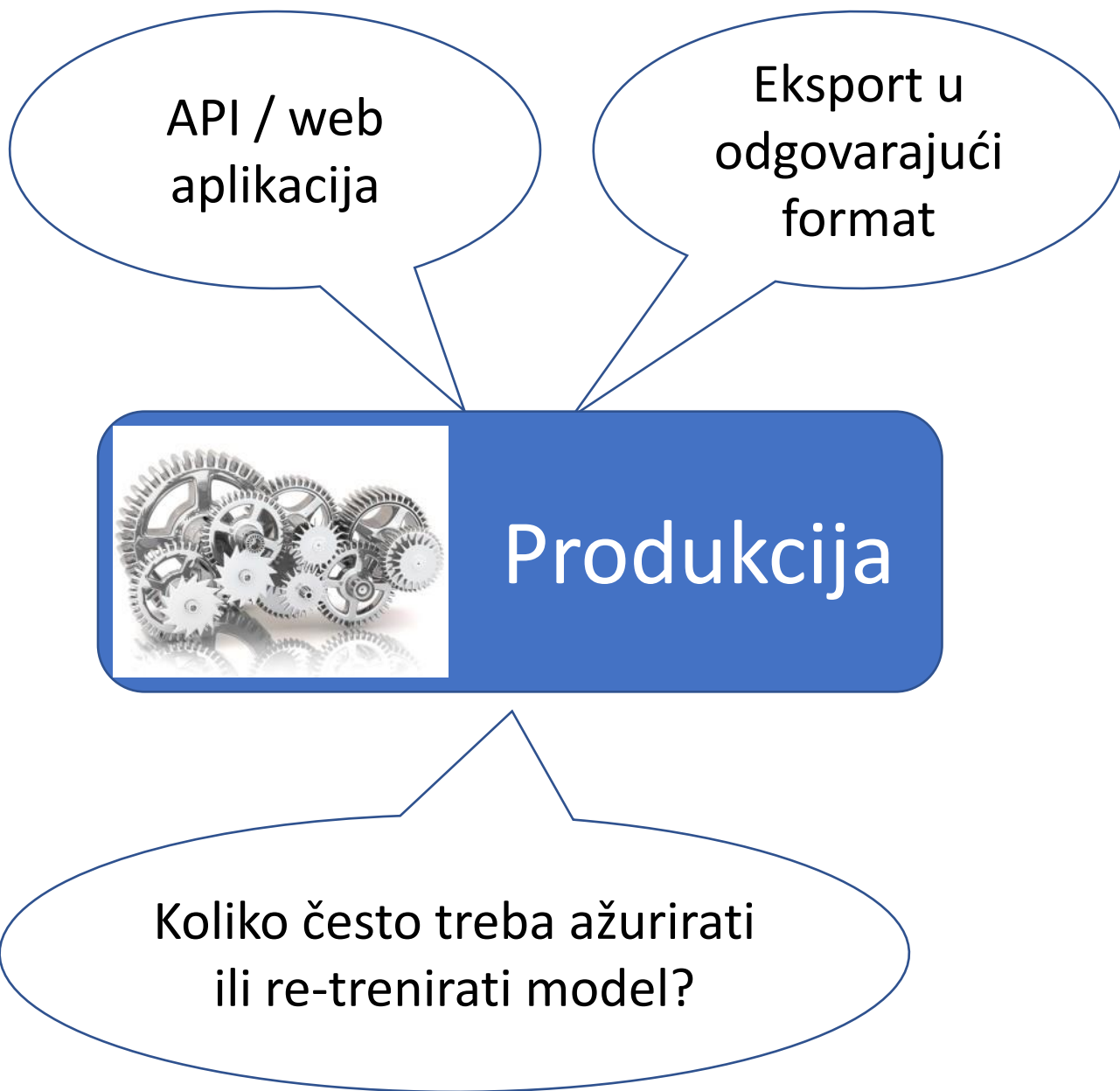


Modelovanje

Odabir
hiperparametara

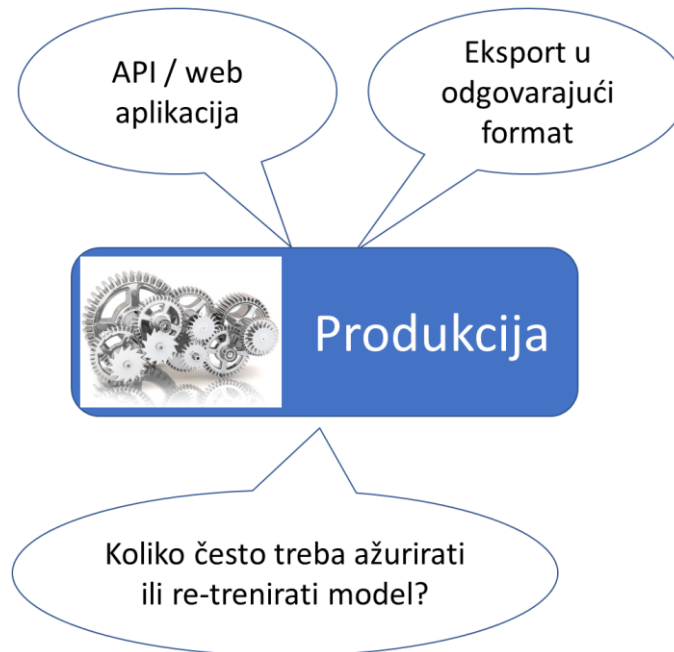
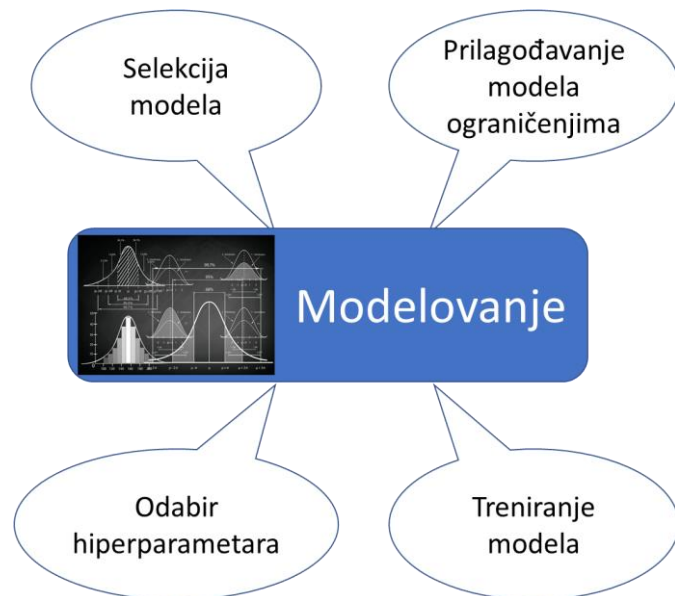
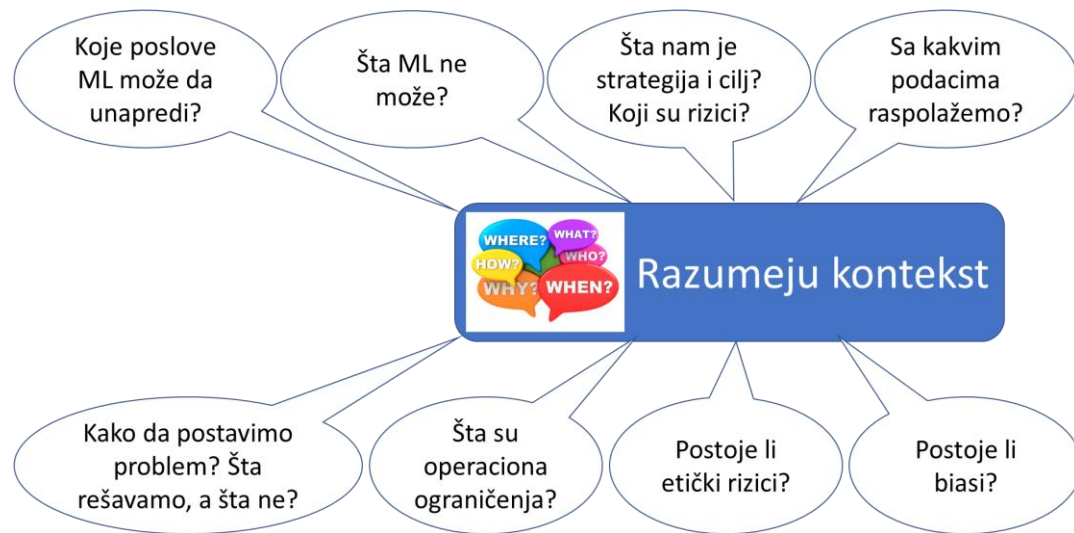
Treniranje
modela

Šta ML
eksperti
zapravo rade?



Šta ML
eksperti
zapravo rade?







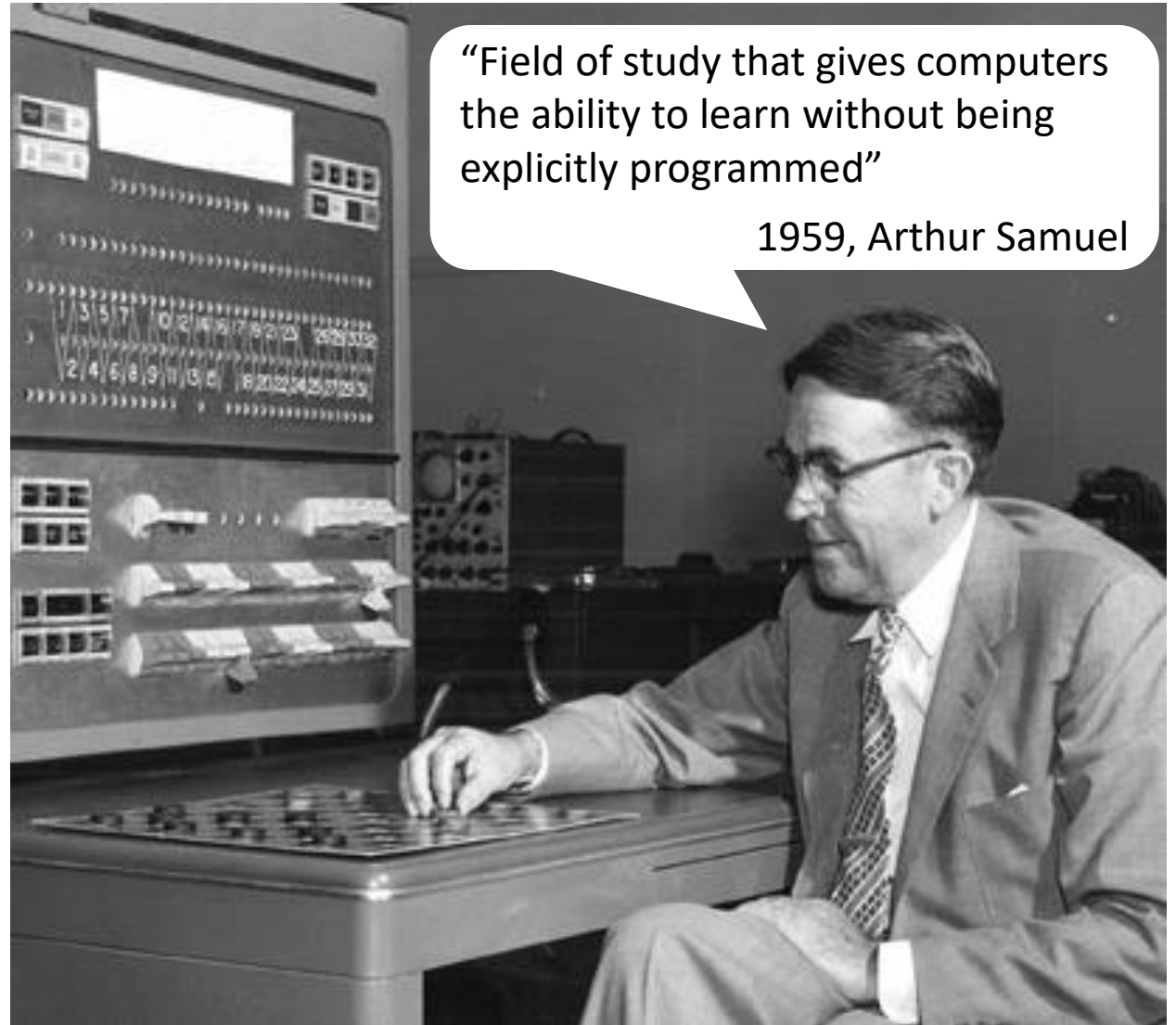
ta želi



ML eksperti su
potrebni

- Kada alati nemaju zadovoljavajuće performanse
- Kada alati ne pružaju prednost nad konkurencijom
- Kada treba razviti nove modele ili tehnike treniranja
- Kada treba interpretirati rezultate
- Kada treba odrediti ulogu mašinskog učenja u organizaciji

Šta je mašinsko učenje?



“Field of study that gives computers the ability to learn without being explicitly programmed”

1959, Arthur Samuel

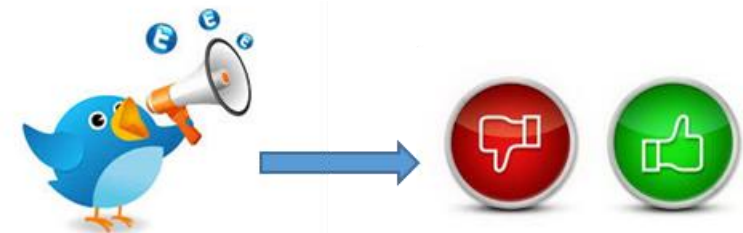
Šta je mašinsko učenje?

A computer program
is said to learn
from experience E
with respect to some
task T and
some performance
measure P ,
if its performance on
 T , as measured by P ,
improves with
experience E .”

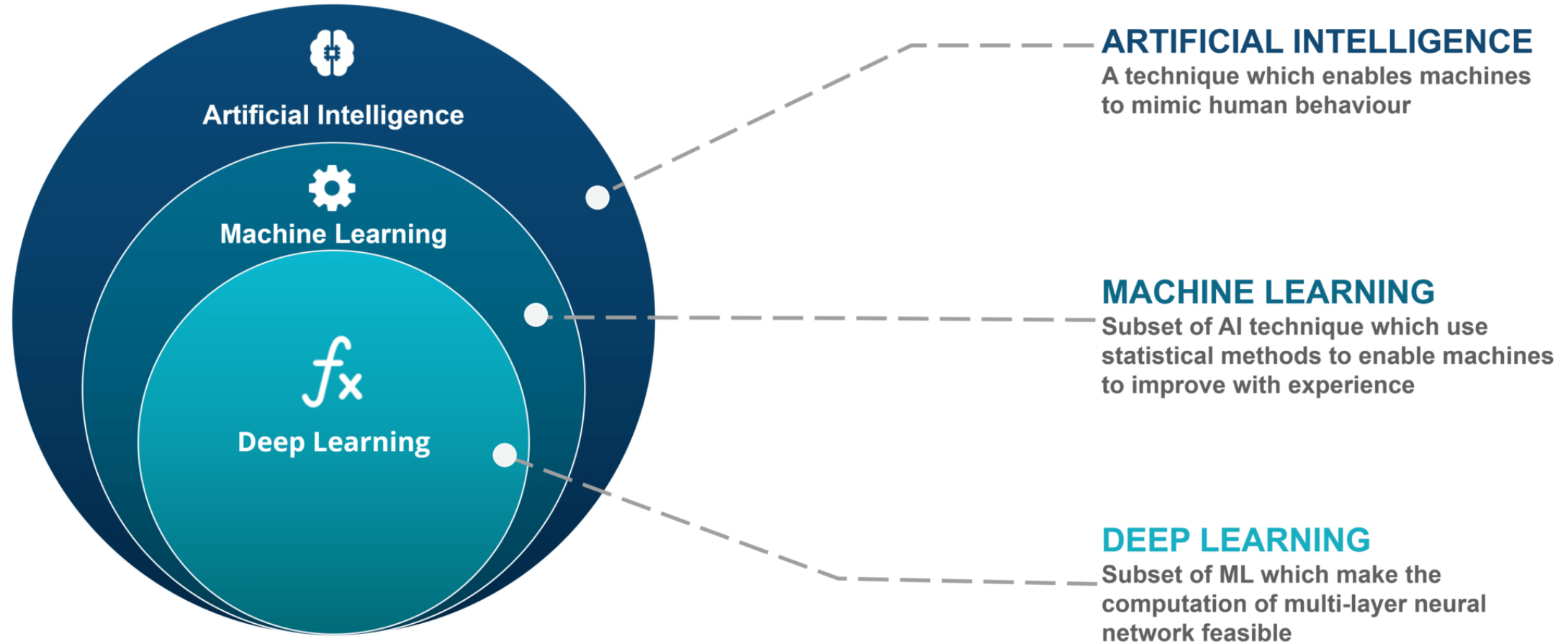
1998, Tom Mitchell



Analizirajte ovu izjavu na primeru
određivanja sentimenta *tweet-a*



Artificial Intelligence vs. Machine Learning vs. Deep Learning



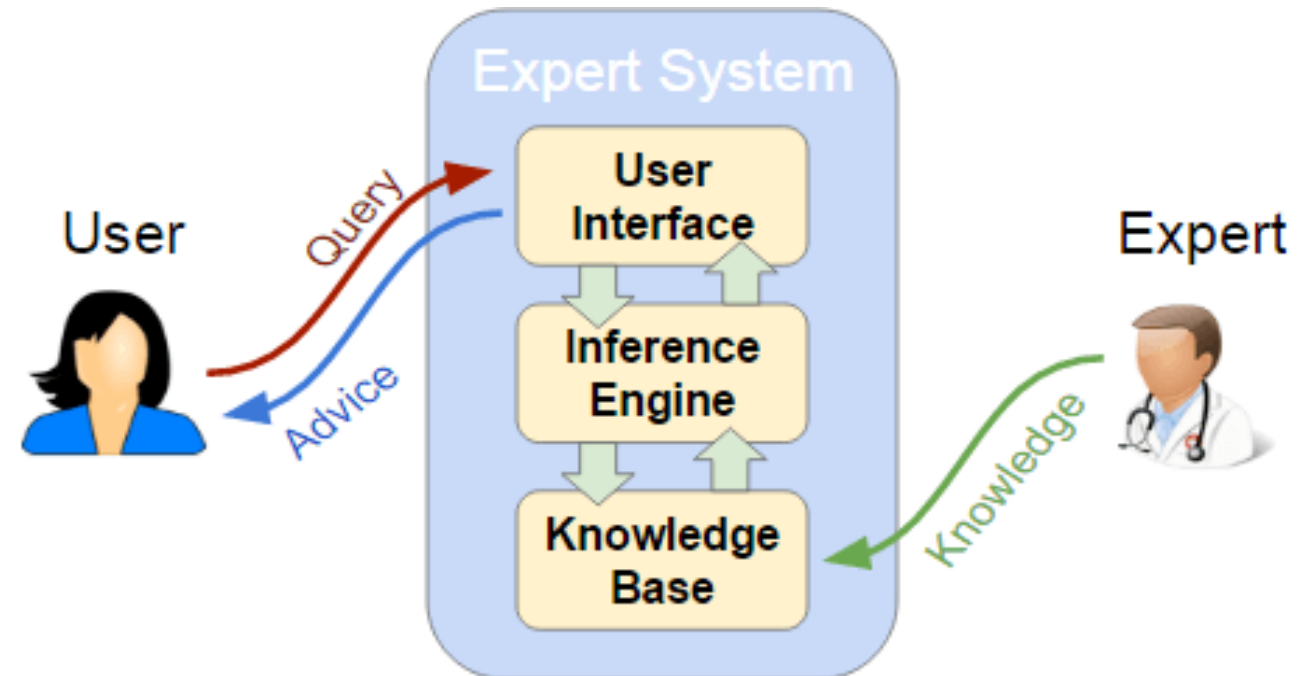
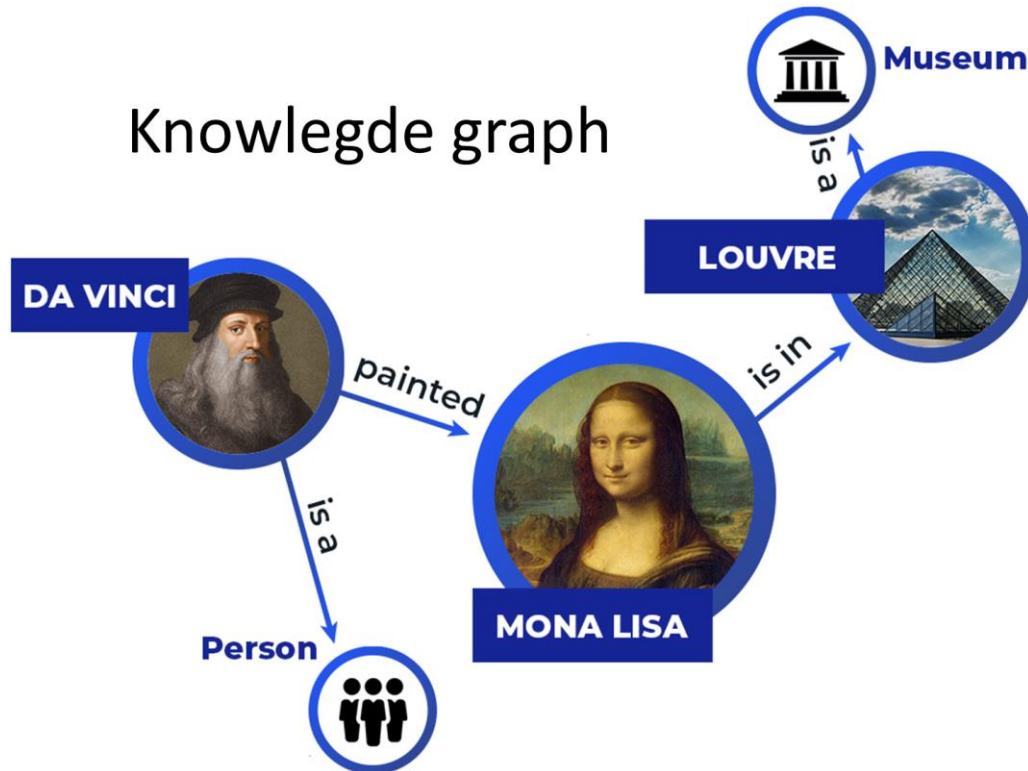
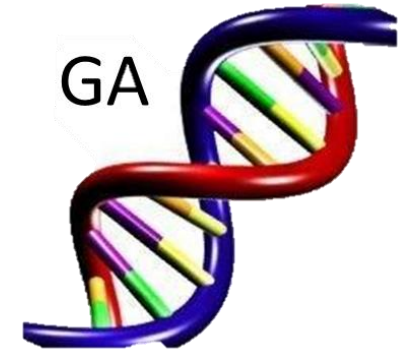
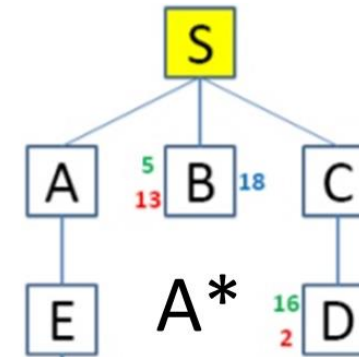
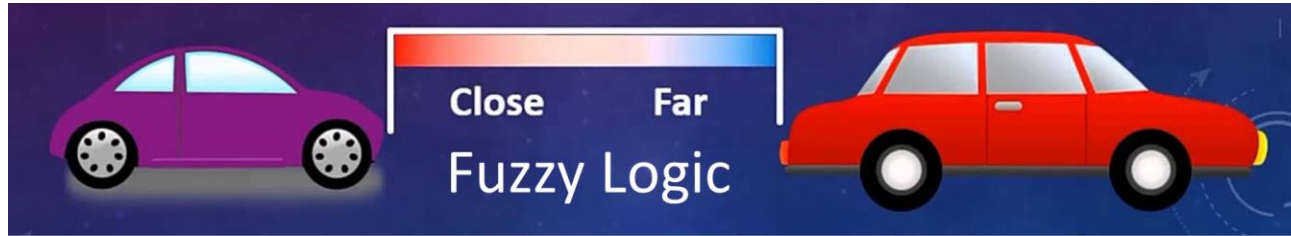
Navedite
primere

(1 min na papiru)

Machine Learning

Good, Old-Fashioned AI
(not ML)

Good, Old-Fashioned AI



Good, Old-Fashioned AI



Da li su danas prevaziđeni?



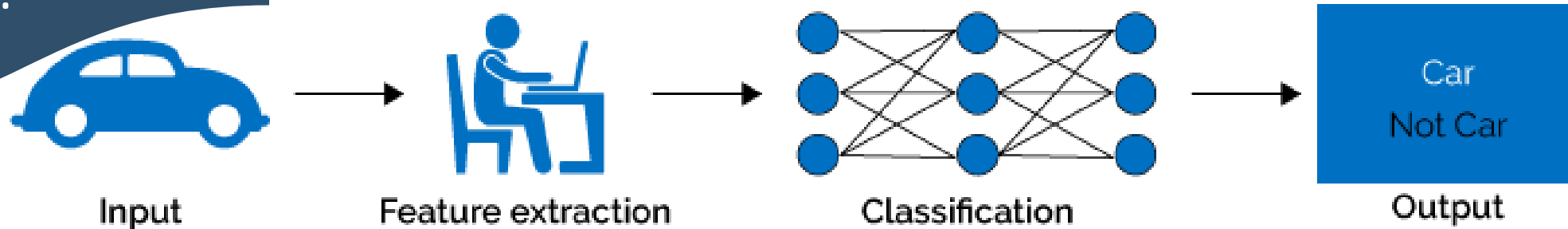
arago

the automation experts

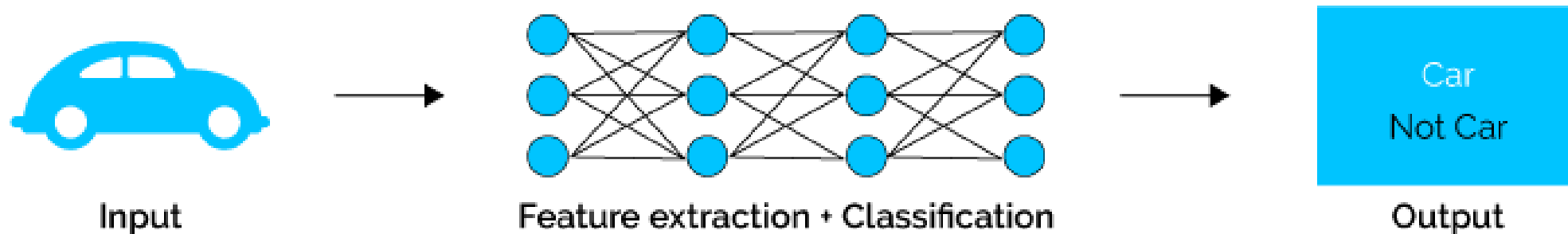
Kombinovan sa ML-om

Šta je ključna
razlika modela
dubokog
učenja i
tradicionalnih
modela?

Machine Learning



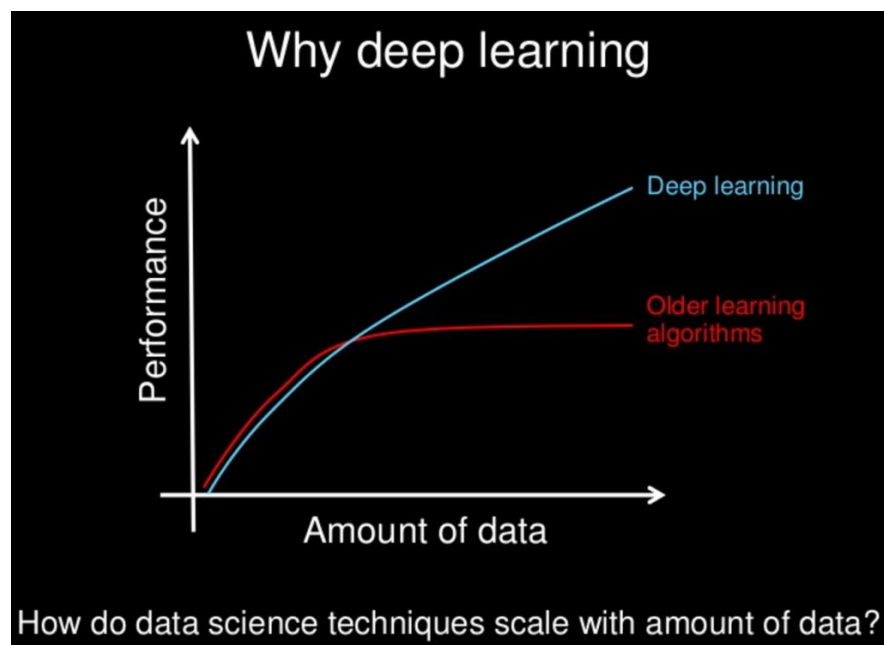
Deep Learning



Upotrebljavaju li se još tradicionalni modeli?

(2 min na papiru): prednosti i mane tradicionalnih i *deep learning* algoritama

	Deep Learning	Tradicionalni modeli
Količina podataka	Veliki skupovi podataka	Mali/srednji skupovi podataka



The analogy to deep learning is that the rocket engine is the deep learning models and the fuel is the huge amounts of data we can feed to these algorithms

Andrew Ng



	Deep Learning	Tradicionalni modeli
Količina podataka	Veliki skupovi podataka	Mali/srednji skupovi podataka
Tip podataka	Nestrukturirani	Strukturirani




Anthony Goldbloom
founder and CEO of [Kaggle](https://www.kaggle.com)

XG Boosting is the engine of choice for structured problems (where feature manufacturing is the key)

However, hyper parameter tuning is only the few % accuracy points improvement on top, the major breakthroughs in predictive power come from feature manufacturing

For unstructured problems (visuals, text, sound) - Neural Networks run the show

	Deep Learning	Tradicionalni modeli
Količina podataka	Veliki skupovi podataka	Mali/srednji skupovi podataka
Tip podataka	Nestrukturirani	Strukturirani
Hardver	Snažne mašine, GPU	Standardan
Ekstrakcija obeležja	Nije potrebna	Potreban domenski ekspert

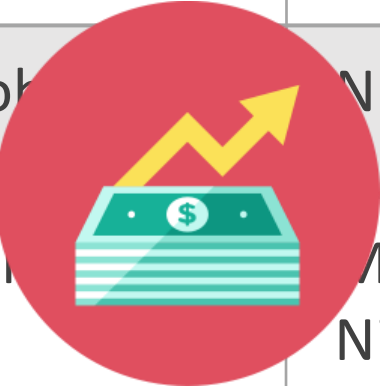
	Deep Learning	Tradicionalni modeli
	<p>Veliki</p> <p>Nestr</p> <p>Snažr</p> <p>Nije p</p>	<p>ka</p>
Vreme treniranja	Može trajati nedeljama. Nije jednostavno	Nekoliko min do nekoliko h

How many of you have devised the deep net from scratch architecture and all and trained it from the ground up and when it didn't work you felt bad about yourself like you did something wrong?

Ali Rahimi

NIPS 2017 Test-of-Time Award presentation

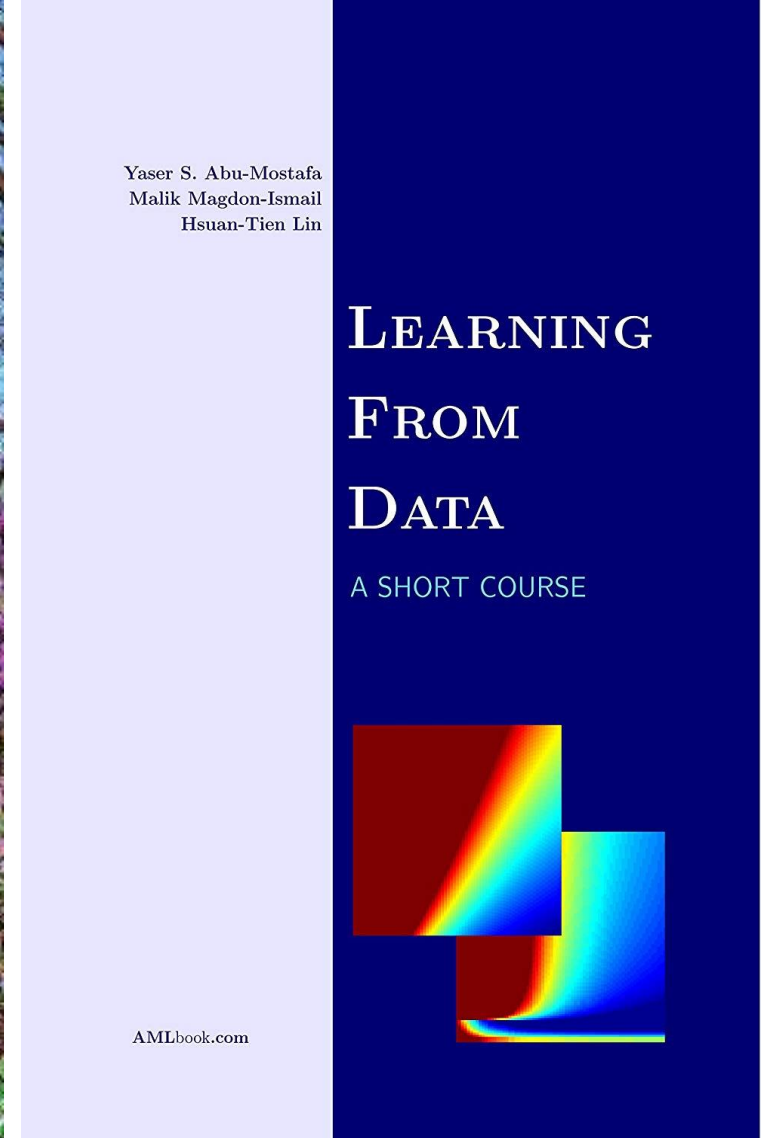
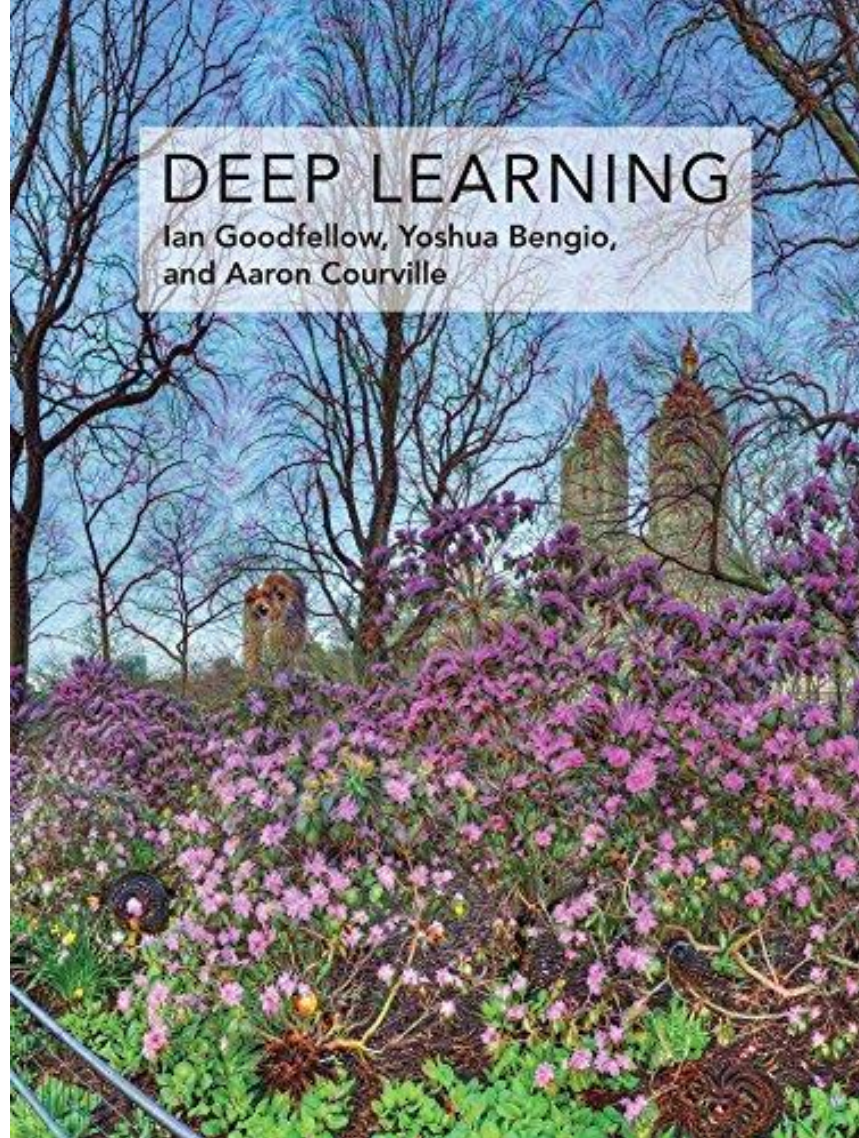
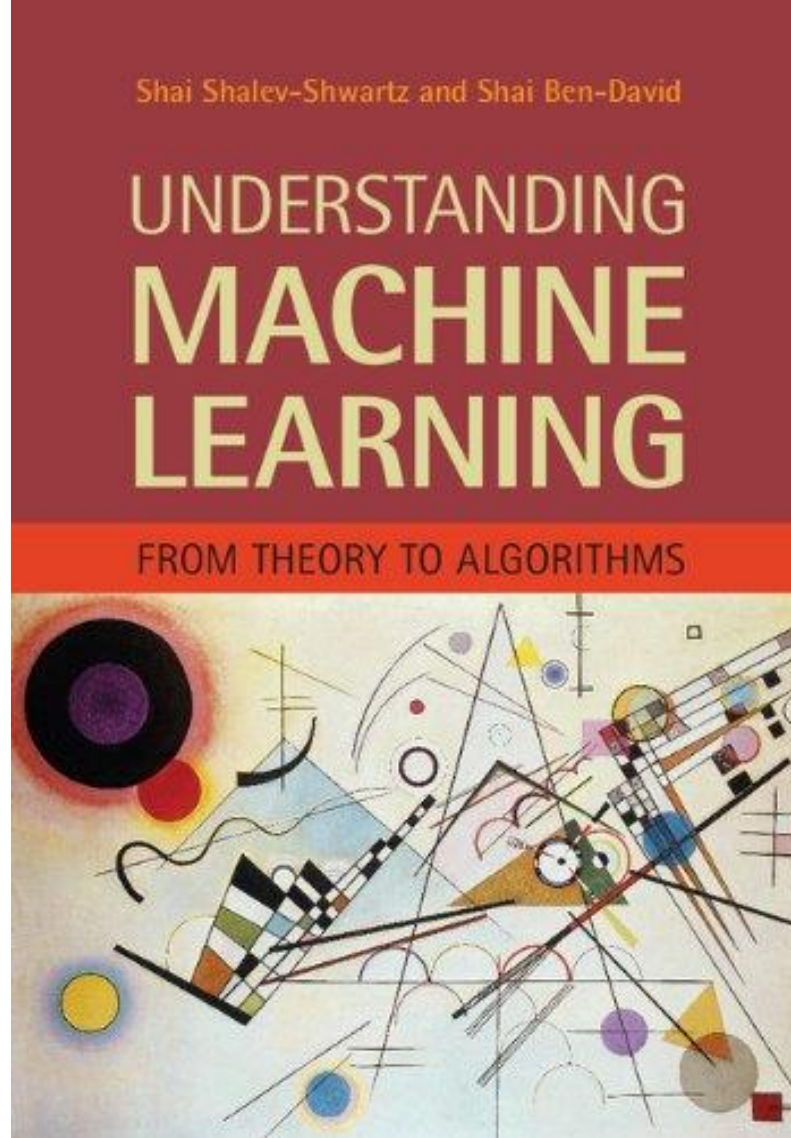
			
Količina podataka		Tip podataka	
Hardver			
Ekstrakcija			
Vreme treniranja			
Interpretabilnost	Teška do nemoguća	Određeni modeli su interpretabilni	



Potrebni su
nam svi alati

GOFAI i ML i DL

- Svaki se koristi u odgovarajućem kontekstu
- Vezivanjem isključivo za jedan alat rizikujemo da zanemarimo nove pravce istraživanja
- Tržištu su zanimljive sve ove veštine



Dodatna literatura: Knjige

Dodatna
literatura

Online kursevi

Machine Learning Specialization

Build Intelligent Applications. Master machine learning fundamentals in four hands-on courses.



4.7

18,254 ratings

UNIVERSITY of WASHINGTON

<https://www.coursera.org/specializations/machine-learning>

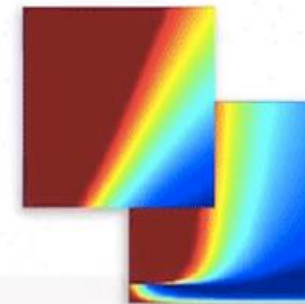


Machine Learning

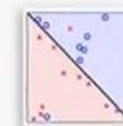
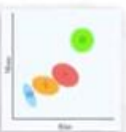
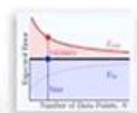
<https://www.coursera.org/learn/machine-learning>

Machine Learning

by Andrew Ng

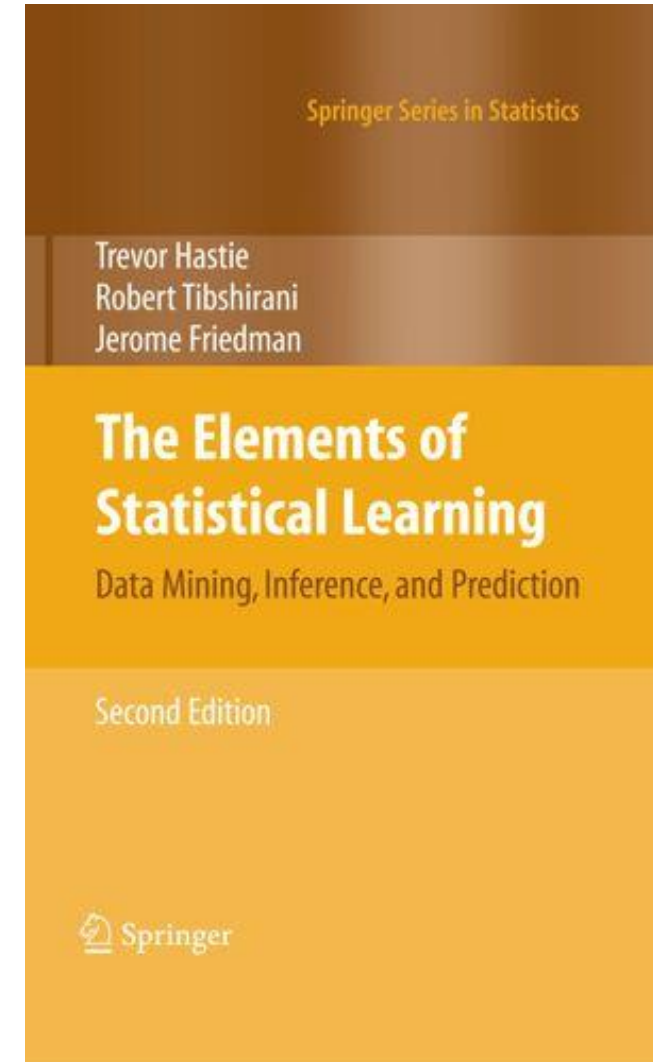
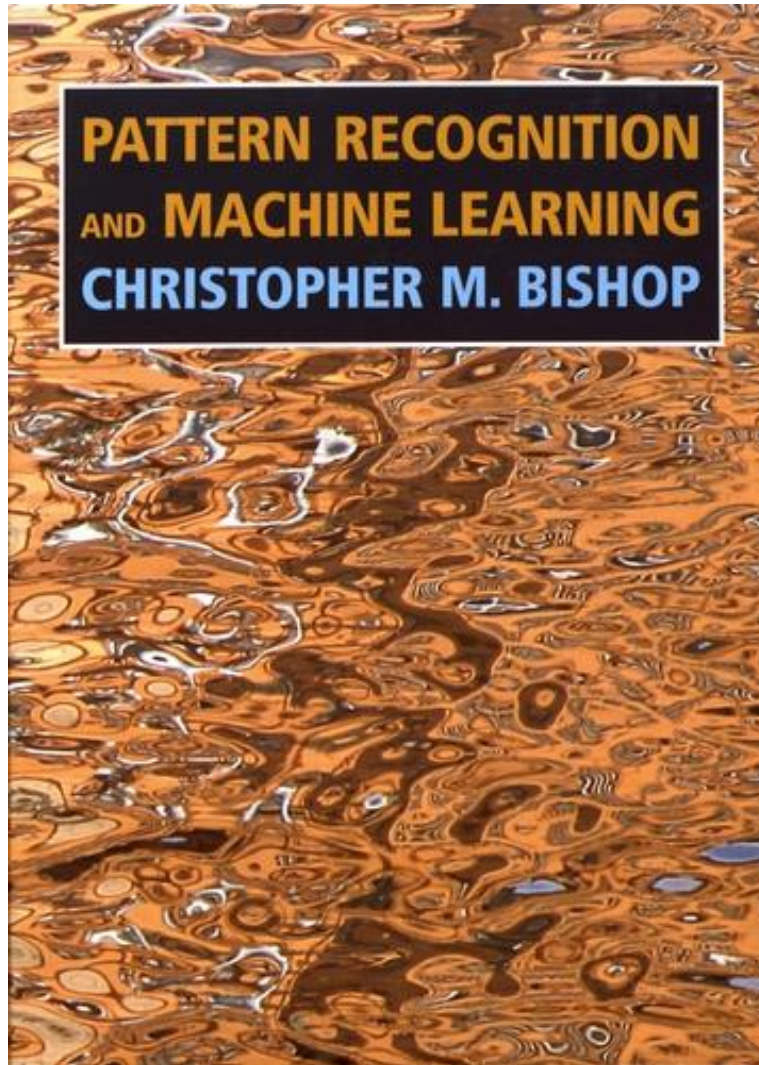


Learning From Data

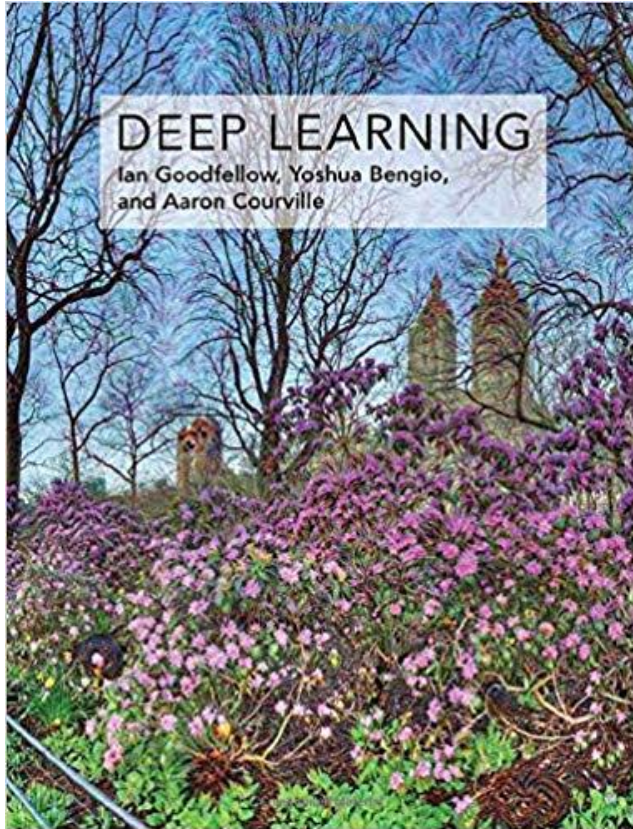


INTRODUCTORY MACHINE LEARNING COURSE

<http://work.caltech.edu/telecourse.html>

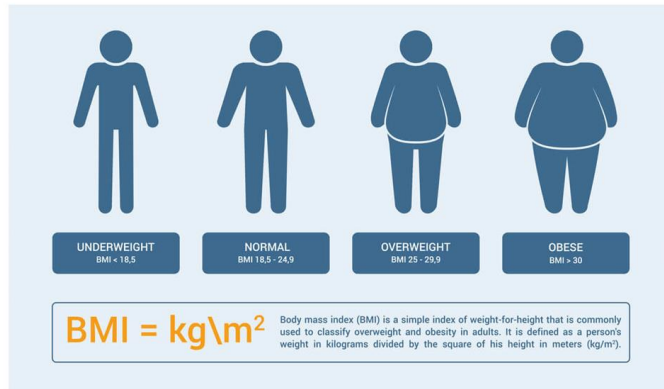


Dodatna literatura: Knjige



Machine learning is essentially a form of applied statistics with increased emphasis on the use of computers to statistically estimate complicated functions and a decreased emphasis on proving confidence intervals around these functions

BODY MASS INDEX

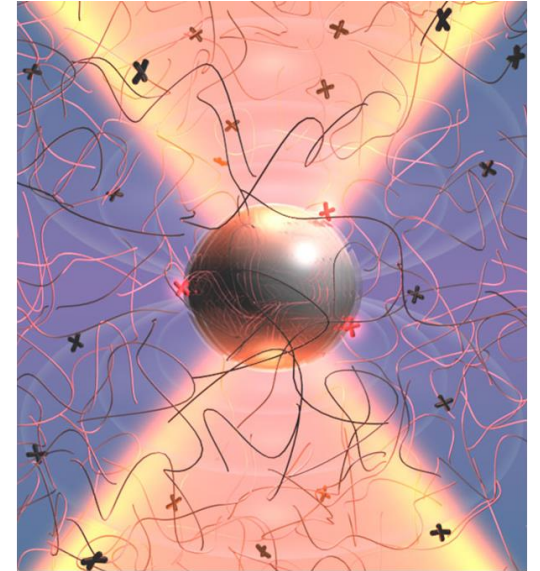


Predikcija BMI osebe na osnovu njene visine i teže

Šta će biti sledeći najpopularniji novi programski jezik?

Ko će pobediti na sledećim izborima?

Ako kreiramo novi proizvod, možemo li predvideti prihod?



Braunovo kretanje – predikcija sledećeg položaja čestice

Ima li smisla rešavati sledeće probleme mašinskim učenjem?

Problem
rešavan
mašinskim
učenjem:

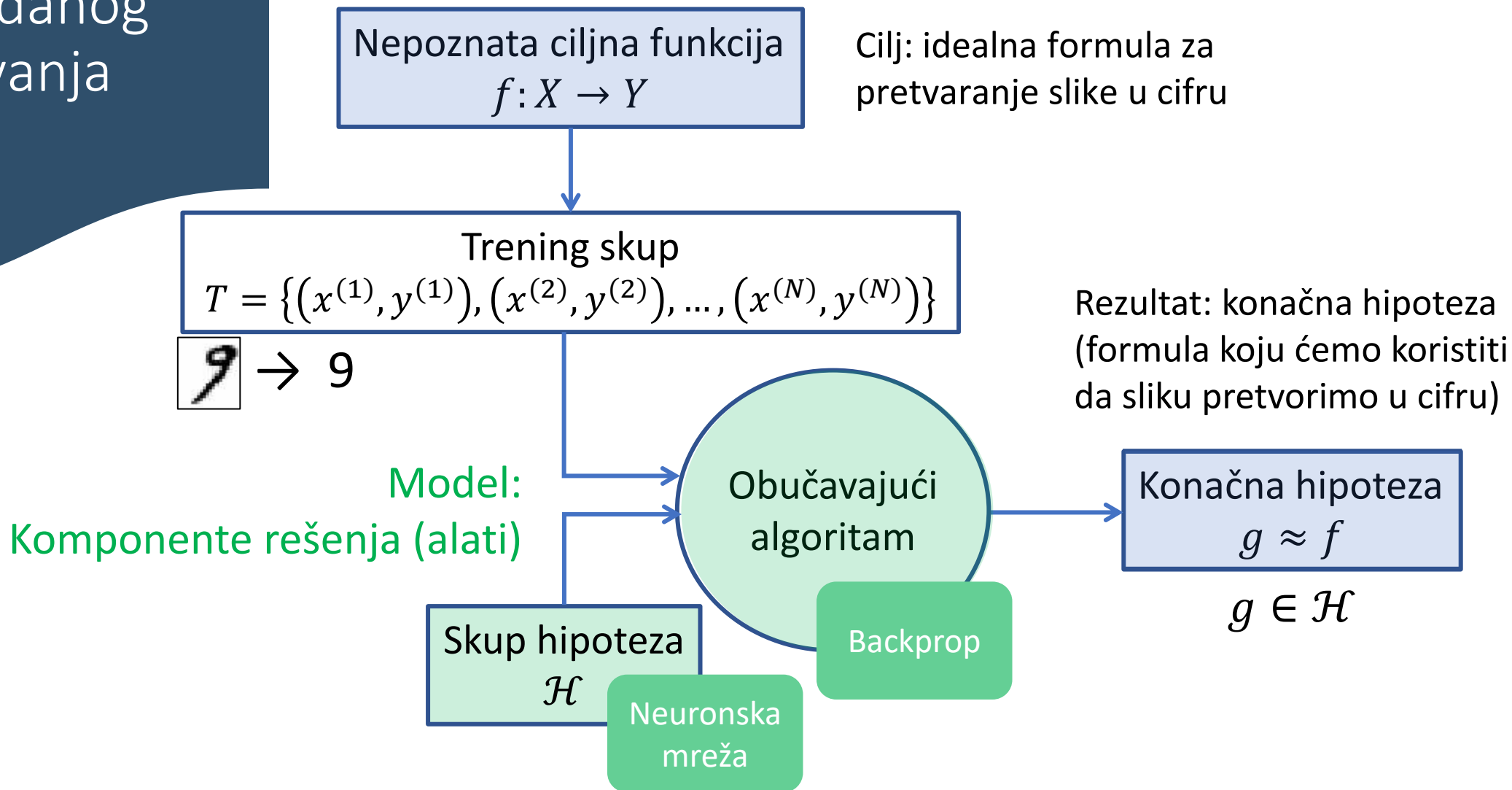
Tri suštinske
komponente

Zadatak: (1 minut na papiru)

Kakve uslove mora zadovoljavati problem da bismo ga mogli (smisleno) rešavati mašinskim učenjem?

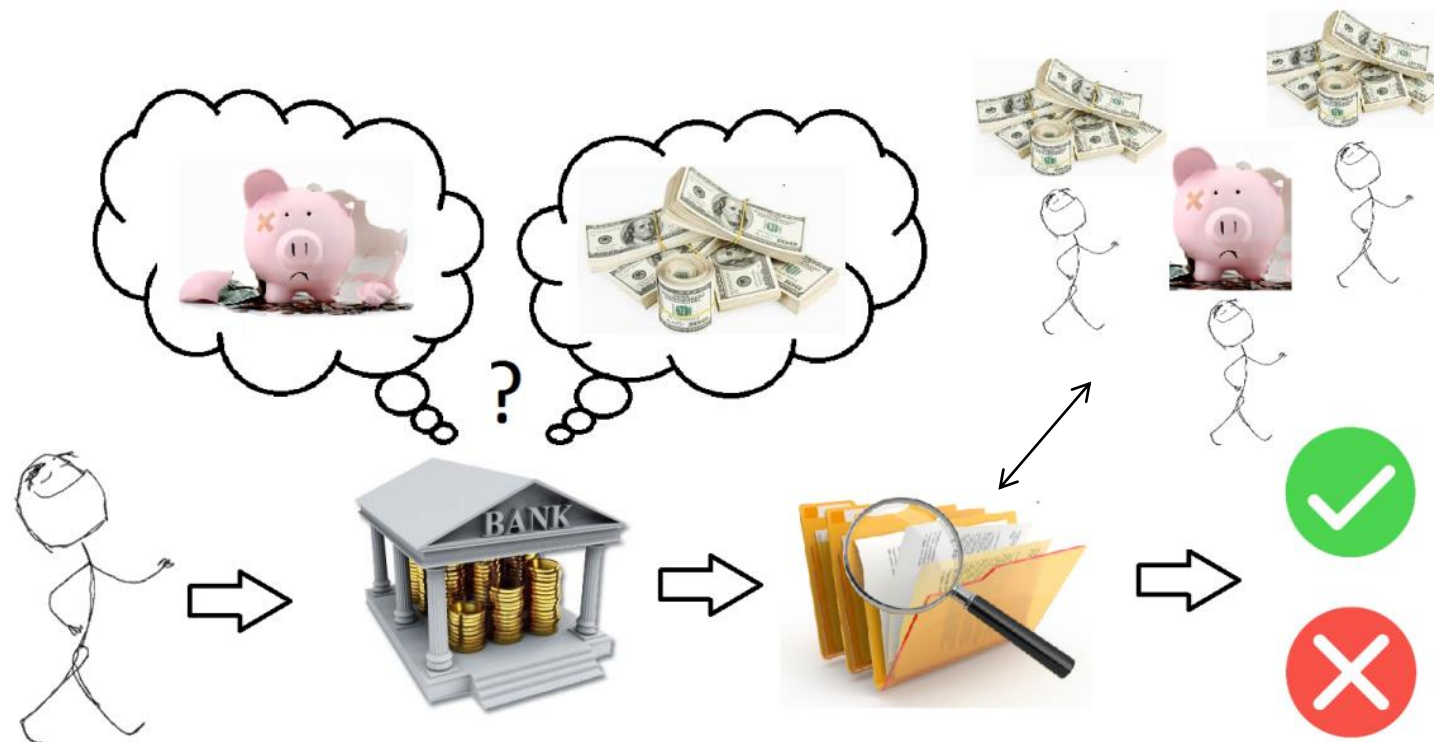
1. Postoji šablon
2. Šablon se ne može izraziti matematičkom zakonitošću
3. Postoje podaci

Komponente nadgledanog obučavanja

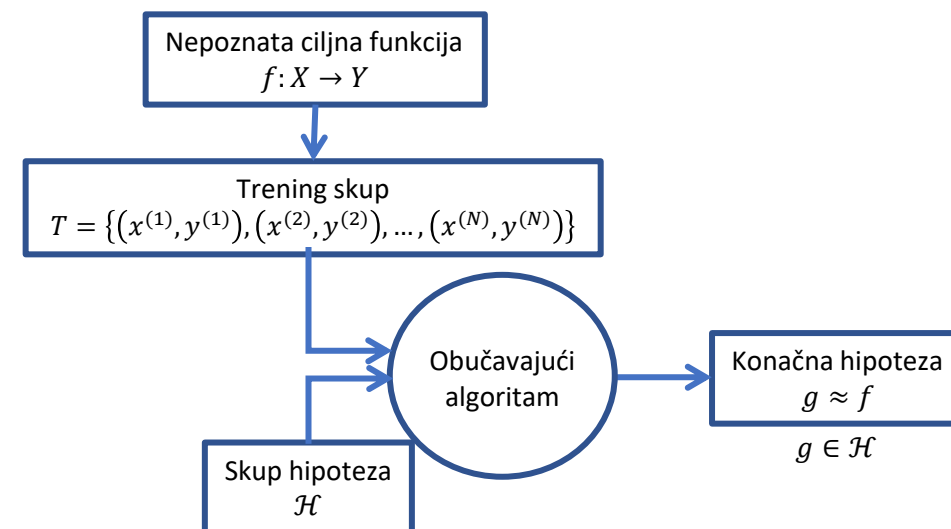


Odobravanje
kredita

odredite
komponente
obučavanja



starost	23 godine
pol	Muški
Godišnja zarada	\$30 000
Trenutni dug	\$15 000
Posедује nekretninu	Da
...	...



starost	23 godine
pol	Muški
Godišnja zarada	\$30 000
Trenutni dug	\$15 000
Posедује nekretninu	Da
...	...

Idealna formula za odobravanje kredita

Ne znamo je, ali se nadamo da je možemo aproksimirati

Nepoznata ciljna funkcija

$$f: X \rightarrow Y$$



Trening skup

$$T = \{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \dots, (x^{(N)}, y^{(N)})\}$$

Istorijski podaci o prethodnim mušterijama

$x^{(i)}$ - podaci iz njihovih molbi za kredit

$y^{(i)}$ - ishod te dodele kredita (da li je banka profitirala ili nije)

Obučavajući
algoritam

Konačna hipoteza

$$g \approx f$$

Skup hipoteza

$$\mathcal{H}$$

Cilj učenja:

Formula koju će banka koristiti da
odredi da li nekome dodeliti kredit

Poznata nam je

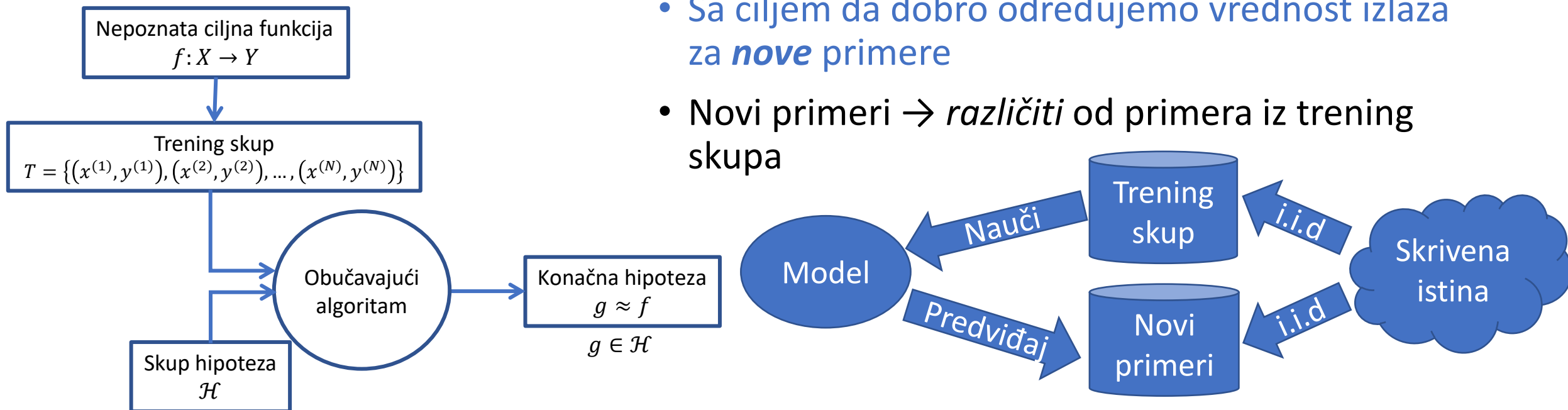
Šta je naš cilj?

Dobra generalizacija

Mapiranje
fotografije na
opisni tekst

Mapiranje
detalja o kući
na njenu cenu

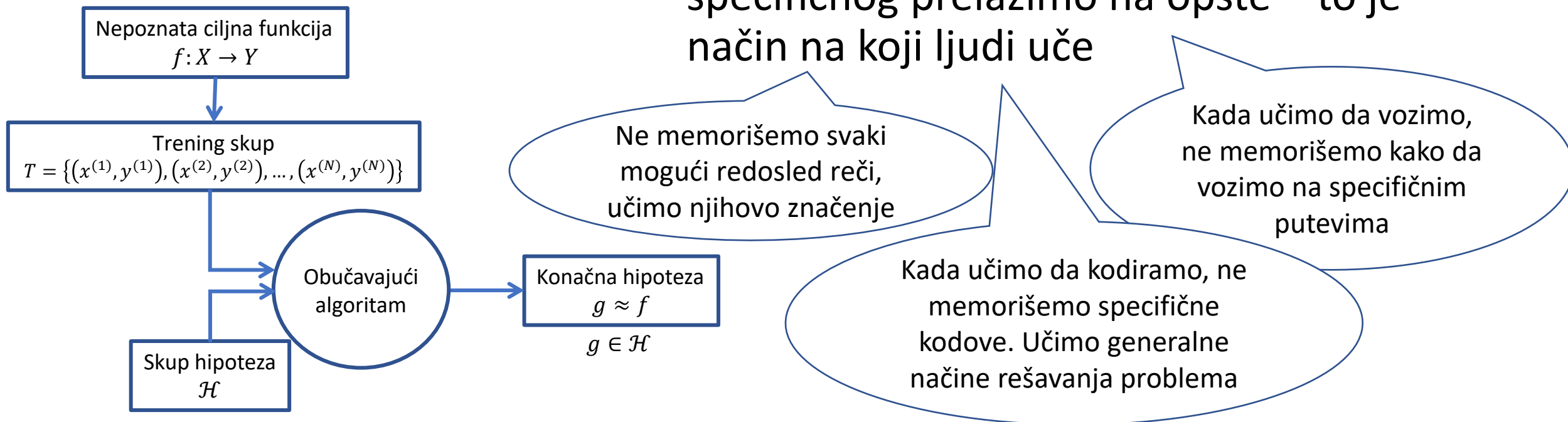
- Mašinsko učenje traži razumnu aproksimaciju mapiranja između šablona u ulaznim podacima i šablona u izlaznim podacima
- Sa ciljem da dobro odredujemo vrednost izlaza za **nove** primere
- Novi primeri → *različiti* od primera iz trening skupa



Šta je naš cilj?

Dobra generalizacija

- Ovo je teško!
- Trening skup moraju biti dobar i pouzdan uzorak
- Da generalizujemo znači da sa specifičnog prelazimo na opšte – to je način na koji ljudi uče



Problem
rešavan
mašinskim
učenjem:

Tri suštinske
komponente

Mašinsko učenje je jedino prikladno kada
problem zahteva generalizaciju

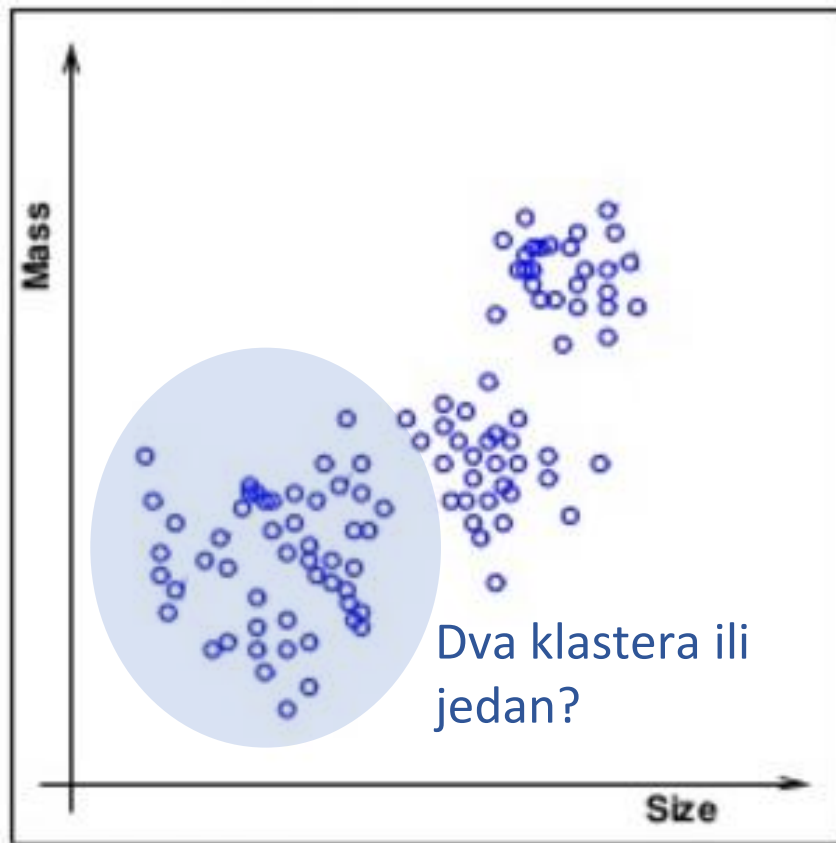
Neki problemi su suviše
kompleksni (berza)

Neki se mogu rešiti
generalizacijom, ali
jednostavnije rešenje
postoji (BMI)

1. Postoji šablon
2. Šablon se ne može izraziti matematičkom zakonitošću
3. Postoje podaci

Neki problemi zahtevaju
preciznost (stanje na
bankovnom računu)

Nenadgledano učenje



Možemo li išta naučiti?

- Možemo podatke podeliti u kategorije iako ne znamo šta kategorije predstavljaju (**klasterovanje**)

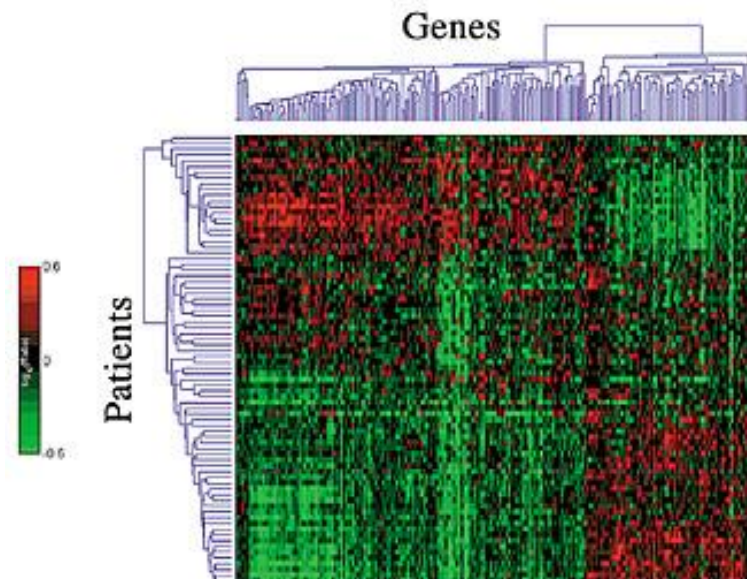
Klasterovanje



Analiza socijalnih mreža



Analiza astronomskih podatka
teorije o rađanju galaksija

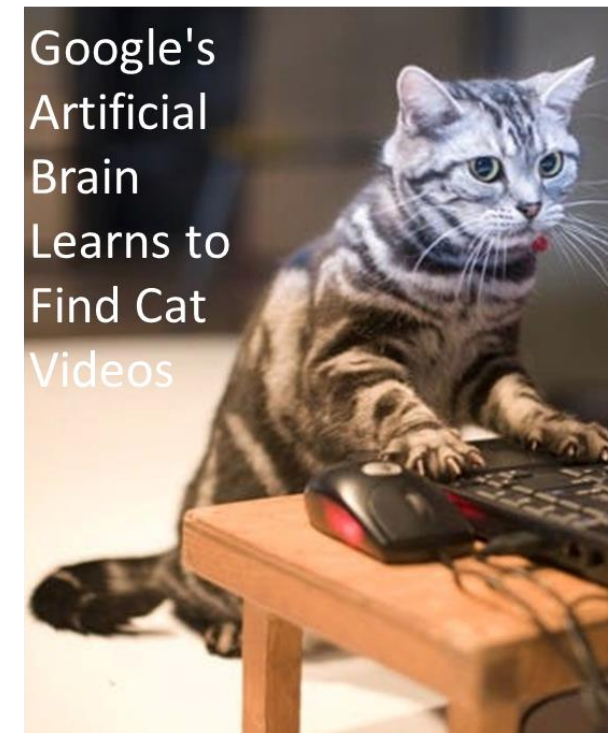


← Gene expression
pattern mostly
associated with
poor outcome

← Gene expression
pattern mostly
associated with
good outcome



Segmentacija mušterija



Google's
Artificial
Brain
Learns to
Find Cat
Videos

Nenadgledano učenje

Zadatak: *pronaći neku strukturu u podacima* (i odrediti obeležja koja su važna za organizaciju podataka)

- Klasterovanje
- Redukcija dimenzionalnosti
- *Density estimation*
- Detekcija anomalija