



N. Bolf\*

Sveučilište u Zagrebu  
Fakultet kemijskog inženjerstva i tehnologije  
Marulićev trg 19, 10 000 Zagreb

## Strojno učenje

Strojno učenje aktualna je tema u istraživanju i u industriji, a nove se metode kontinuirano pojavljuju i razvijaju. Algoritmi strojnog učenja "uče" informacije i odnose među njima izravno iz podataka ne oslanjajući se na teorijske jednadžbe i matematičke modele. Pri tome se modeli s vremenom unaprjeđuju porastom broja uzoraka dostupnih za učenje. Brzina razvoja i složenost algoritama otežavaju praćenje novih tehnika čak i stručnjacima, a pogotovo početnicima. Ovim prilogom upoznat ćemo se sa strojnim učenjem i osnovnim pojmovima.

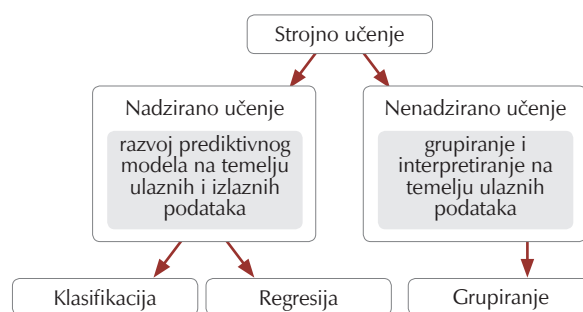
Zadatak algoritma strojnog učenja je pronaći prirodne uzorke i poveznice u podacima te na temelju toga steći uvid i zatim odlučiti i predviđati. Primjenjuju se već svakodnevno za donošenje bitnih odluka u medicinskoj dijagnostici, trgovanju i mešetarenju dionicama, predviđanju potrošnje energije itd. Web-pretraživači mnogih portala oslanjaju se na strojno učenje kako bi nam iz milijuna mogućnosti "prosiljale" i preporučile npr. skladbu, film ili srodnu dušu. Trgovci ga već naširoko primjenjuju da bi stekli uvid u navike svojih kupaca.

Neprestanim porastom količine podataka koji se kontinuirano stvaraju i kolaju svijetom strojno učenje postalo je posebno važno na više područja. Nabrojimo neka:

- **Računalne financije** – za procjenu vaše kreditne sposobnosti kad vam treba dodijeliti kredit ili potporu, kao i za algoritamsku prodaju;
- **Obrada slika i računalna vizija** – kod prepoznavanja lica, detekcije gibanja i detekcije raznih objekata;
- **Računalna biologija** – detekcija tumora, istraživanje lijekova i sekvenciranje DNA;
- **Proizvodnja energije** – prognoza opterećenja elektroenergetske mreže i kretanja cijena;
- **Industrijska proizvodnja** – u prediktivnom održavanju, procjeni i nadziranju kvalitete proizvoda;
- **Obrada prirodnog jezika** – raspoznavanje govora, razumijevanje i generiranje prirodnog jezika.

Osnovnu podjelu, slika 1, možemo svesti na:

- **Nadzirano učenje** (engl. *supervised learning*)  
Modeli se uče na temelju poznatih ulaznih i izlaznih podataka prikupljenih eksperimentom ili akvizicijom tako da mogu predviđati buduće izlaze na temelju dostupnih ulaza;
- **Nenadzirano učenje** (engl. *unsupervised learning*)  
Zadatak nenadziranog učenja je pronaći skrivene uzorke i inherentne strukture u ulaznim podacima bez poznavanja izlaza.



Slika 1 – Osnovna podjela strojnog učenja

### Nadzirano učenje

U ovom načinu učenja algoritmi rade s poznatim skupom ulaznih podataka (**ulazi** modela) i poznatim skupom izlaznih podataka (**izlazi** modela) i uvježbavaju model za predviđanje (predikciju).

Nadzirano učenje primjenu nalazi u sljedećim postupcima:

- **Klasifikacija ili razvrstavanje** – Predviđanje **diskretnih odziva**, pri čemu se ulazni podatci razvrstavaju u kategorije. Primjeri su je li e-pošta originalna ili neželjena (*spam*), kod medicinske dijagnostike: je li tumor karcinogen ili dobroćudan, prepoznavanje riječi i fraza kod govora, prepoznavanje znakova i drugih zapisa i sl.
- **Regresija** – Predviđanje **kontinuiranih odziva**. Regresijom se predviđaju kontinuirane varijable, poput promjene temperature ili koncentracije u reaktoru, sastav i kvaliteta proizvoda, fluktuacije potrošnje električne energije u elektroenergetskom sustavu i sl.

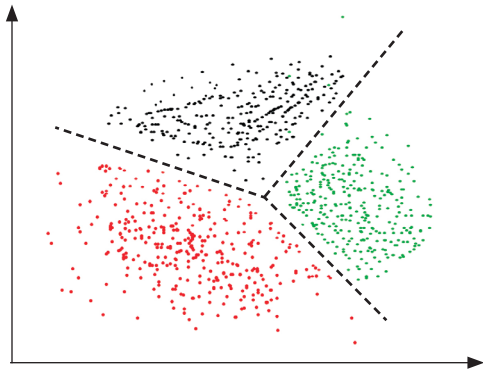
### Nadzirano učenje za predviđanje infarkta

Pretpostavimo da kliničar želi predvidjeti hoće li pacijent u dogledno vrijeme doživjeti srčani udar. Podatci o pacijentu poznati specijalistu su njegova dob, težina, visina i krvni tlak. U arhivi klinike postoje informacije o povijesti bolesti dosadašnjih pacijenata (npr. tipični simptomi prije srčanog udara). Sada je zadatak algoritma kombinirati postojeće podatke u model koji će predviđati je li novi pacijent u skoro vrijeme u opasnosti od srčanog udara i shodno tome prepisati terapiju.

\* Prof. dr. sc. Nenad Bolf  
e-pošta: [bolf@fkit.hr](mailto:bolf@fkit.hr)

### Nenadzirano učenje

Nenadzirano učenje pronalazi skrivene uzorke ili intrinzične strukture u podacima. Njime se **grupiraju uzorci** ili **otkrivaju strukture**, slika 2. Primjenjuju se za izvođenje zaključaka iz skupa podataka koje čine samo ulazni podatci bez poznavanja odzivnih podataka (izlaza). Najčešća tehnika nenadziiranog učenja je **grupiranje** (engl. *clustering*), pri čemu se traže skrivene obrasci ili grupe. Primjenu nalazi pri analizi sekvenca gena, analizi tržišta i prepoznavanju objekta.



**Slika 2** – K-means algoritam najjednostavniji je algoritam grupiranja. Razdjeljuje  $n$  opservacija u  $k$  grupa, pri čemu svaki uzorak pripada grupi čijem je prototipu (u ovom slučaju srednjoj vrijednosti) najbliži.<sup>1,2</sup>

### Kako odabrati algoritam?

Izbor algoritma je velik, a danas postoji više desetaka algoritama. Ukratko, svaki od algoritama ima drugačiji pristup učenju. Mogli bismo reći da ne postoji najbolja niti univerzalna metoda. Čak i iskusni eksperti ne mogu predvidjeti hoće li algoritam biti djelotvoran pa preostaje metoda pokušaja i pogreške. Odabir algoritma ovisi o **broju** i **vrsti** podataka, što se želi postići te gdje i kako će se primijeniti rezultati. Na slici 3 dana je podjela i popis tehnika, dok će zainteresirani čitatelj više informacija o algoritmima naći u literaturi.

### Kada primijeniti strojno učenje?

Strojno učenje treba primijeniti kod rješavanja **kompleksnih zadataka** ili problema koji uključuju veliku količinu podataka i velik broj varijabli, ali ne postoje razrađene formule ili jednadžbe (modeli). Evo nekoliko primjera:

- Klasični algoritmi ili pravila previše su složeni za prepoznavanje ljudskog lica ili interpretaciju govora;
- Kad se pravila koja definiraju zadatak konstanto mijenjaju, kao npr. kod detekcije prijevara iz postojeće baze financijskih transakcija bankovnih računa;
- Kad je priroda podatka takva da se stalno mijenja i program se tome mora prilagoditi – npr. automatizirana prodaja, predviđanje potreba za energijom i predviđanje trendova u prodaji.

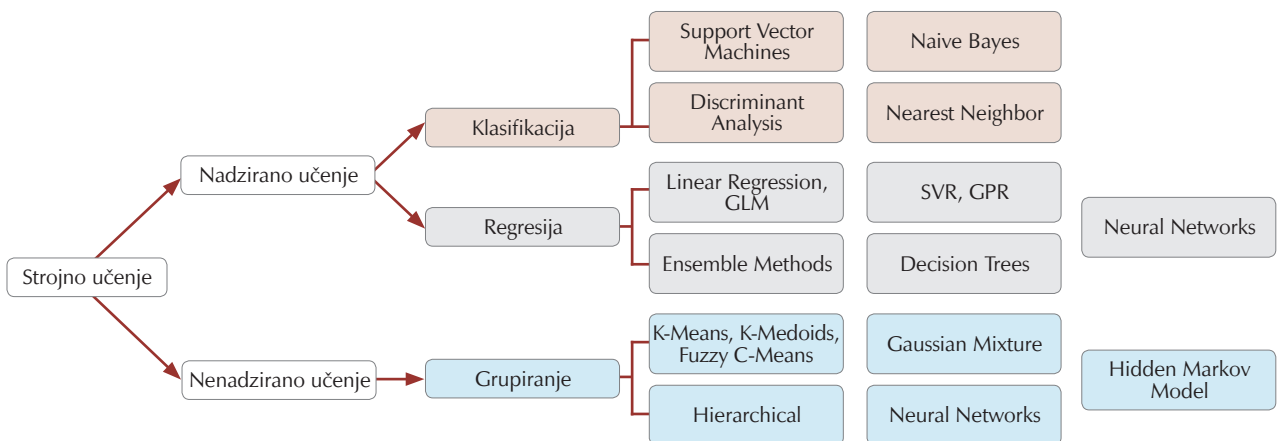
### Primjer analize umjetničkih djela

Evo jednog zanimljivog primjera u kojem se strojno učenje pokazalo na djelu.

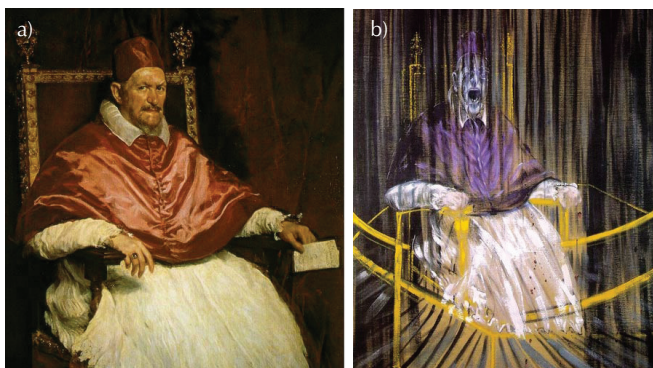
Istraživači *Laboratorija za umjetnost i umjetnu inteligenciju* na Sveučilištu Rutgers željeli su ustanoviti može li računalni algoritam klasificirati slike prema stilu, žanru i umjetniku. Započeli su prepoznavanjem vizualnih značajki za klasificiranje stila slikara. Algoritmi koje su razvili razvrstavaju stilove slika u bazi podataka sa 60 % točnosti, što nadmašuje tipične laike. Istraživači su pretpostavili da su vizualne značajke korisne za klasifikaciju stila (što spada u **nadzirano učenje**) također se mogu upotrijebiti za određivanje umjetničkih utjecaja (to je **nenadzirano učenje**). Primijenili su algoritme za klasifikaciju koji su učili na slikama dostupnim putem Googleove tražilice. Algoritme su ispitali na više od 1700 slika od 66 različitih umjetnika koji su stvarali u rasponu od 550 godina.

I što se desilo!? Algoritam je lako prepoznao stilski povezana djela, npr. utjecaj *Portreta pape Inocenta X* (završenog oko 1650.), slika 4a, španjolskog slikara **Diega Velazqueza** na *Studiju Velazquezovog portreta pape Inocenta X* (iz 1953.), slika 4b, engleskog slikara irskog podrijetla **Francisa Bacona** istaknutog predstavnika nove figuracije.

U tom slučaju ključni korak u postupku klasifikacije jest preslikavanje informacija o karakteristikama slikarskog djela kao što su kombinacija boja, modulacija svjetlosti i sjene, mekoća rubova, jedinstveni slikarski rukopis sadržan u potezima kista, materičnosti namaza boje i karakterima linija, u skup numeričkih veličina dobivenih primjenom raznih metoda za ekstrakciju značajki slike.<sup>3</sup>



**Slika 3** – Podjela metoda strojnog učenja (tehnike provedbe su napisane engleskim jezikom zbog lakšeg snalaženja u literaturi!)



**Slika 4** – a) Portret pape Inocenta X. naslikan je oko 1650. tehnikom ulje na platnu. Nalazi se u Galeriji Dorije Pamphili u Rimu;<sup>4</sup> b) Studija nakon Velázquezova Portreta pape Inocenta X iz 1953. Nalazi se u Umjetničkom centru Des Moines u Iowi, SAD.<sup>5</sup>

### Primjer primjene za štednju energije

Uvjerili smo se ponekad, posebno u prijelaznim razdobljima godišnjih doba, da sustavi za grijanje, ventilaciju i hlađenje u poslovnim zgradama, bolnicama i velikim trgovačkim centrima nisu djelotvorni. Razlog tome je što ne uzimaju u obzir nagle i sezonske promjene vremena, varijabilne troškove električne energije i toplinska svojstva zgrada. **Inteligentna “cloud” platforma** rješava taj problem. Platforma primjenjuje napredne algoritme i strojno učenje za kontinuiranu obradu informacija s električnih brojila, mjerila temperature i tlaka, isto kao i vremenske prognoze i cijene energije. Strojno učenje segmentira podatke i određuje relativni doprinos potrošnje plina, struje, pare i solarne energije na proces grijanja i hlađenja. Tijekom normalnog rada takvi sustavi smanjuju potrošnju energije HVAC sustava u velikim objektima od 10 % do 25 %.<sup>6,7</sup>



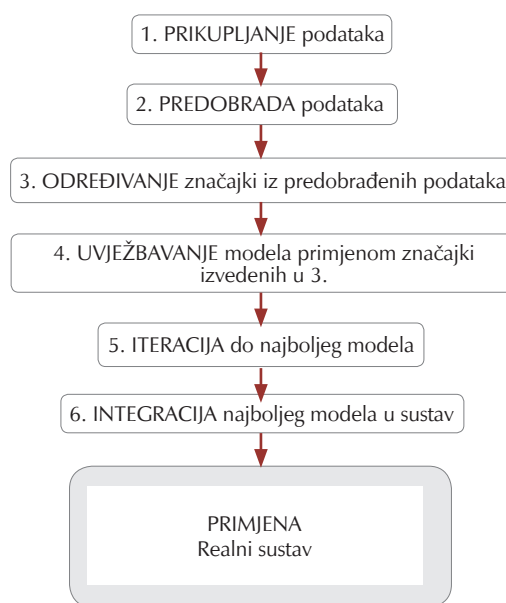
### Tehnika rada sa strojnim učenjem

Kod razvoja modela strojnog učenja rijetkokada će sve ići glatko od početka do kraja. Obično se iteracijski pokušavaju različiti pristupi i ideje. Većina izazova vezana je uz baratanje podacima i pronalazak pravog modela. Datoteke s podacima javljaju se u mnogo različitih oblika i veličina. Podatci iz stvarnih sustava mogu biti nestrukturirani, nepotpuni i u različitim formatima. Ponekad su to samo jednostavni brojevi. No često je to kombinacija neko-

liko različitih vrsta podataka, (npr. signali sa senzora, tekst, slika s kamera i sl.). Predobrada podataka obično zahtijeva posebna znanja i alate. Tako npr. da bismo odabrali karakteristike algoritma za detekciju objekta, potrebna su specijalizirana znanja o obradi slike. U analizi trendova procesnih varijabli i spektrometrijskih podataka potrebna su znanja iz kemijskog inženjerstva i dinamike procesa, kao i analize signala. U svakom slučaju, tri ključna pitanja koja trebate razmotriti su:

- S kojom vrstom podataka radite?
- Koje spoznaje želite izvući iz podataka?
- Gdje i kako će se te spoznaje primijeniti?

Više o tome bit će, nadajmo se, prilike pisati u jednoj od rubrika ili u preglednom radu.



**Slika 5** – Koraci/faze pri razvoju modela primjenom strojnog učenja

### Literatura

1. URL: <https://www.mathworks.com/solutions/machine-learning.html> (pristup: 7. 9. 2021.).
2. URL: <https://www.geeksforgeeks.org/clustering-in-machine-learning/> (pristup: 7. 9. 2021.).
3. URL: <http://www.zemris.fer.hr/predmeti/kdisc/Sem2.pdf> (pristup: 7. 9. 2021.).
4. URL: [https://hr.wikipedia.org/wiki/Portret\\_pape\\_Inocenta\\_X](https://hr.wikipedia.org/wiki/Portret_pape_Inocenta_X) (pristup: 7. 9. 2021.).
5. URL: [https://hr2.wiki/wiki/Study\\_after\\_Vel%C3%A1zquez%27s\\_Portrait\\_of\\_Pope\\_Innocent\\_X](https://hr2.wiki/wiki/Study_after_Vel%C3%A1zquez%27s_Portrait_of_Pope_Innocent_X) (pristup: 7. 9. 2021.).
6. URL: <https://www.brainboxai.com/> (pristup: 7. 9. 2021.).
7. URL: [https://www.researchgate.net/publication/340839625\\_Air\\_Conditioning\\_Energy\\_Saving\\_from\\_Cloud-Based\\_Artificial\\_Intelligence\\_Case\\_Study\\_of\\_a\\_Split-Type\\_Air\\_Conditioner](https://www.researchgate.net/publication/340839625_Air_Conditioning_Energy_Saving_from_Cloud-Based_Artificial_Intelligence_Case_Study_of_a_Split-Type_Air_Conditioner) (pristup: 7. 9. 2021.).
8. URL: <https://www.adriasecuritysummit.com/smart-hvac-systems-on-cybersecurity-and-connection-protocols/> (pristup: 7. 9. 2021.).