



Ponovitev osnov strojnega učenja

Kako se ekspert odloča?



- Je pacient bolan ali zdrav?
 - Odgovor: Kategorija (bolan/zdrav)
- Kaj bi naredil ekspert (zdravnik)?
 - Pregledal bi kartoteko pacienta (težo, višino, krvno sliko, hormonsko sliko, temperaturo telesa, krvni pritisk...).
 - Odločil bi se na podlagi znanja (ki jih je pridobil iz informacij, izkušenj in dognanj).
- V glavi ima zgrajen **model znanja**. Primer:
 - Če je telesna temperature pacienta nad 40°C, je pacient bolan.
 - Če je telesna temperature nad 38°C in je sistolični krvni pritisk nad 140 mHg, je pacient bolan.
 - Če je telesna temperature pod 38°C, je pacient zdrav.

Prezentacija modela znanja

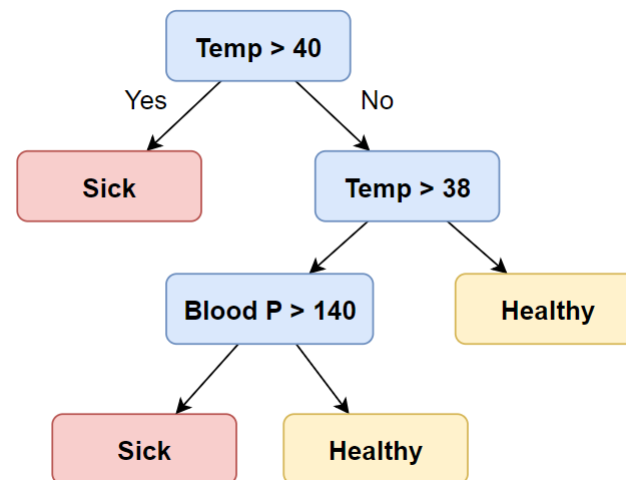


- **Model znanja** (angl. *knowledge model*) je formalen zapis znanja. Na papir lahko model znanja zapišemo na različne načine (v različnih podatkovnih strukturah).

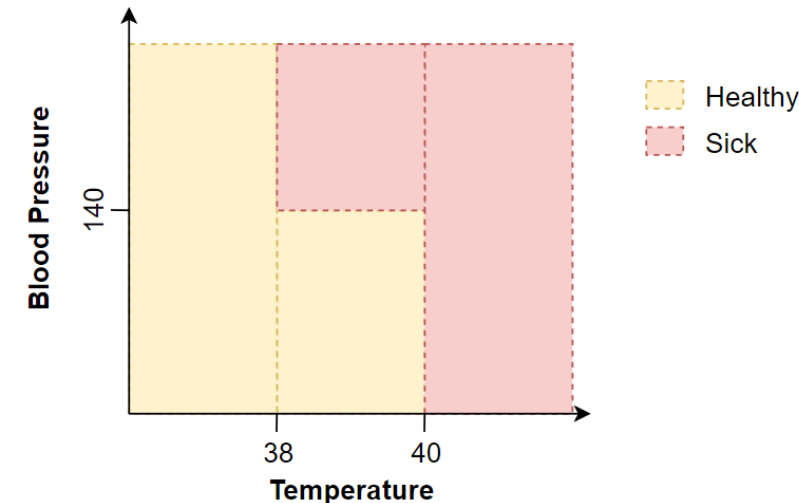
Pravila

- Če je telesna temperature pacienta nad 40°C , je **pacient bolan**.
- Če je telesna temperature nad 38°C in je sistolični krvni pritisk nad 140 mHg, je **pacient bolan**.
- Če je telesna temperature pod 38°C , je **pacient zdrav**.
- ...

Odločitvena drevesa



Matematične funkcije

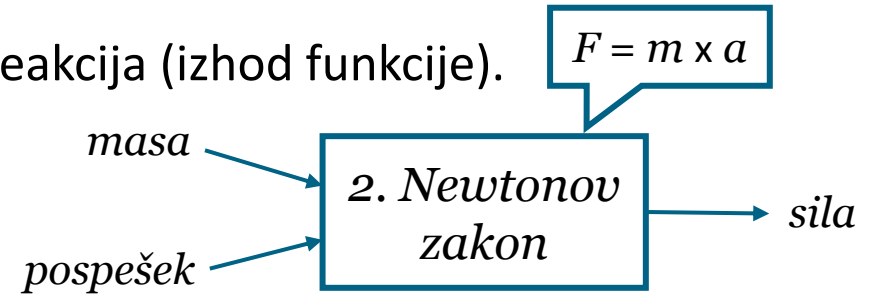


Model znanja – formalno

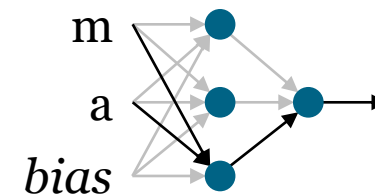
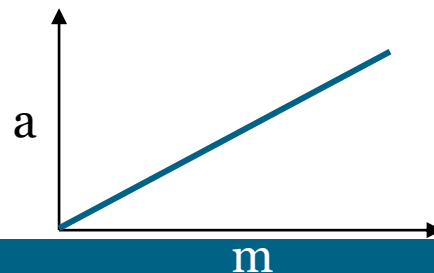
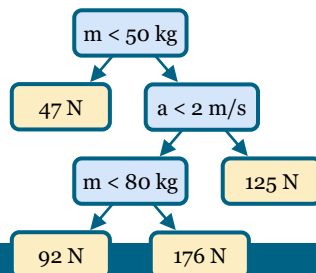


- **Funkcije** v realnosti (in ne računalništvu) so matematični izrazi s katerimi opisujemo **pojave v naravi oz. realnosti**.

- Glede na stanje (vhodne informacije) se izvede neka reakcija (izhod funkcije).



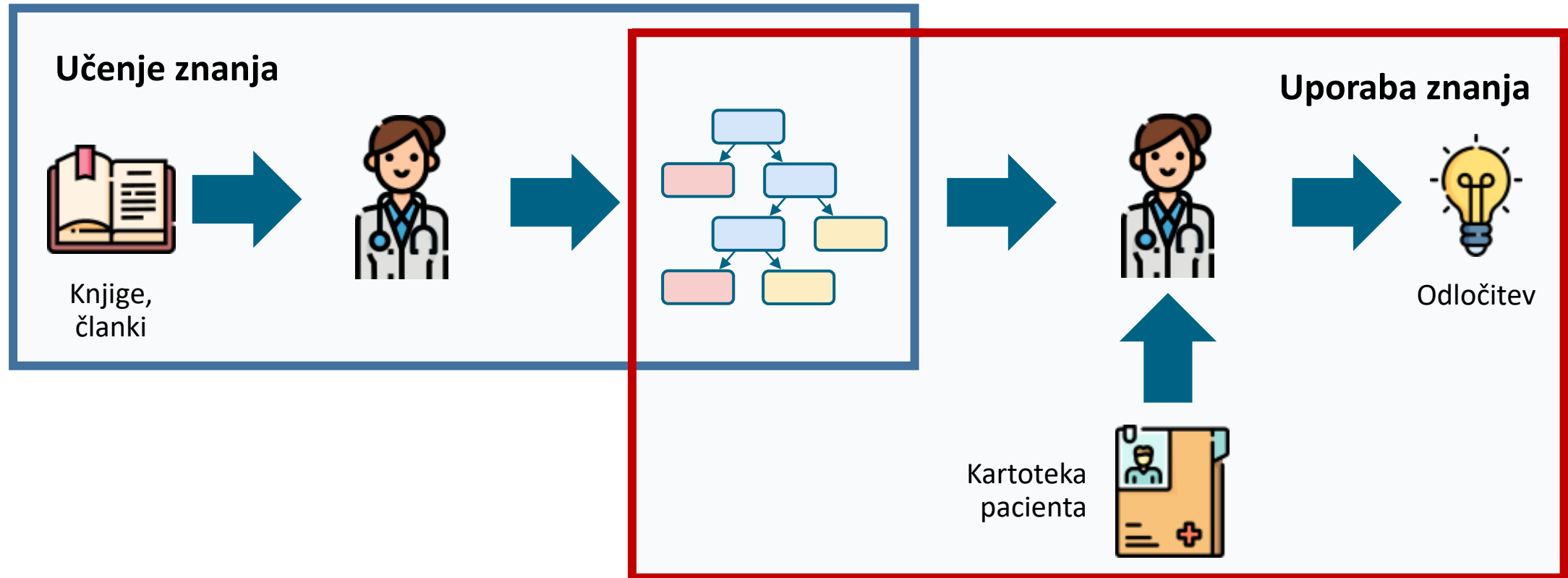
- **Model** je (poenostavljena) predstavitev funkcije (pojava iz realnosti).
 - Zakaj poenostavljena? Največkrat ne poznamo vseh sestavnih delov funkcije in vseh vhodnih podatkov, zato vemo le v grobem opisati pojave.
 - Primeri: ekonomski modeli, fizikalni model, biološki model človeškega telesa, *model znanja*...



Uporaba modela znanja



- Modele znanje (iz papirja ali iz možganov) eksperti uporabljajo naprej pri svojem delu.

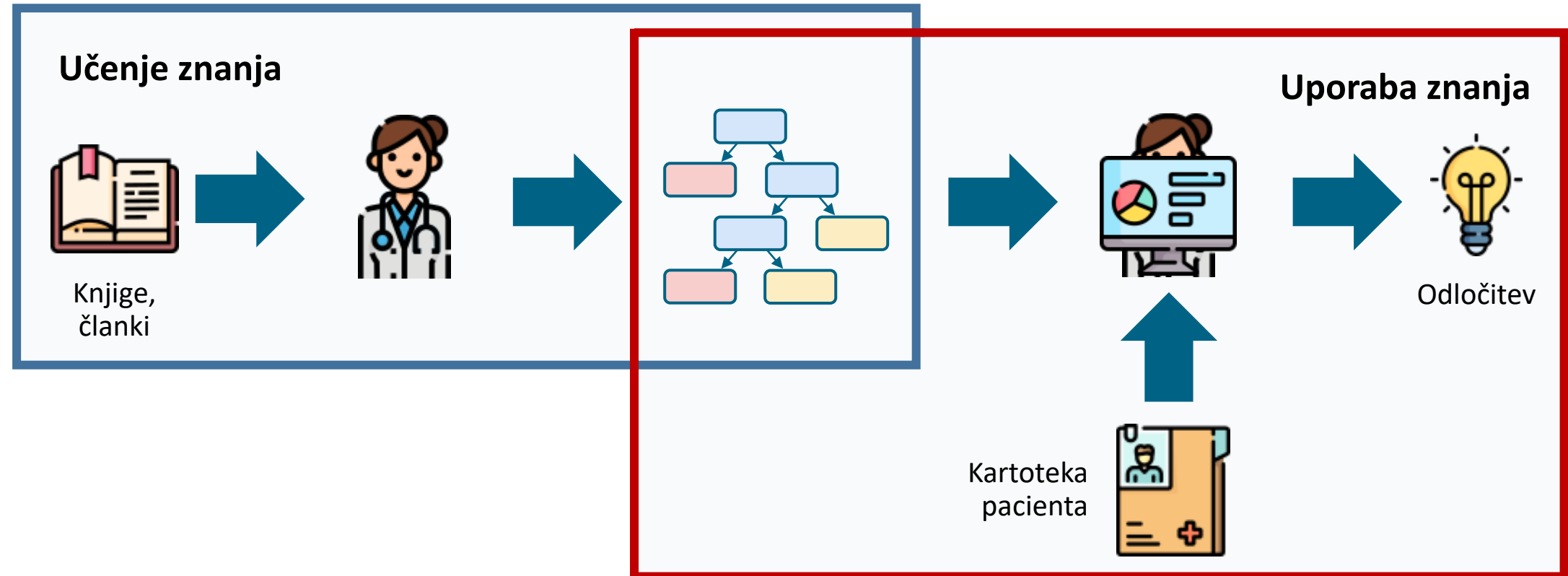


Ekspertni sistem

Ekspertni sistem se odloča sam s pomočjo modela znanja, ki ga je zgradil ekspert.



- Računalnik lahko prebere modele znanja in se na podlagi njih odloča.
 - Informatiki lahko prepišejo model znanja v programsko kodo.

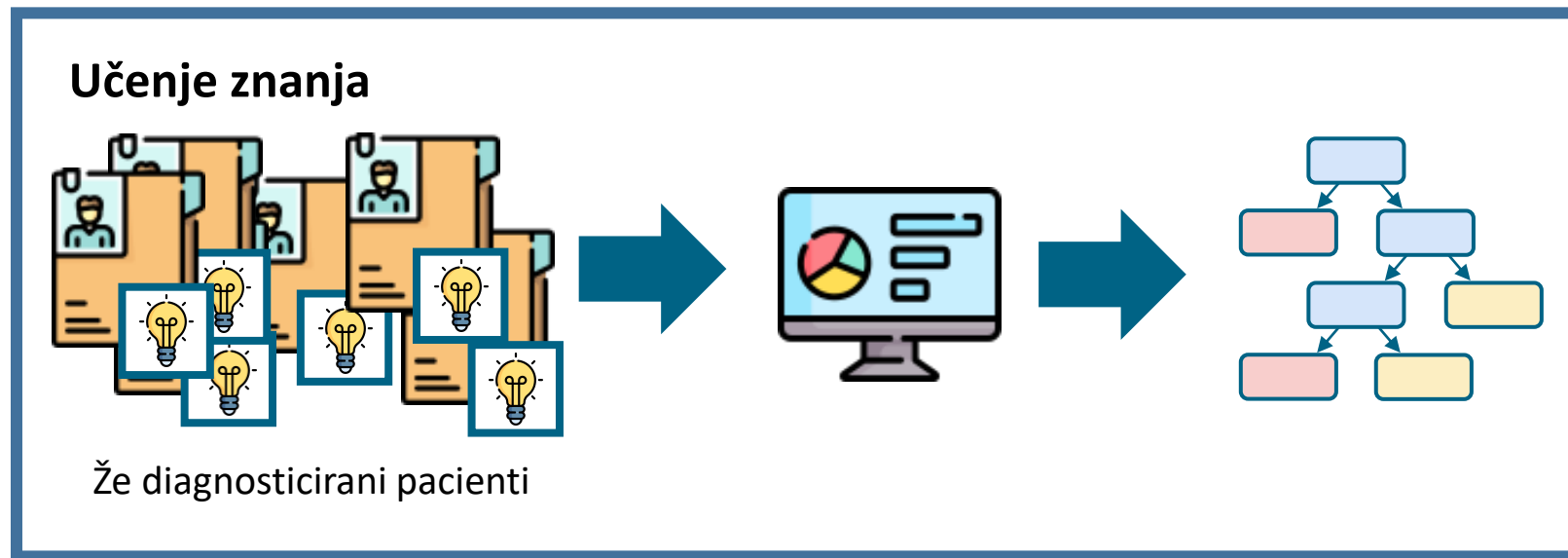




Kako se stroj uči?



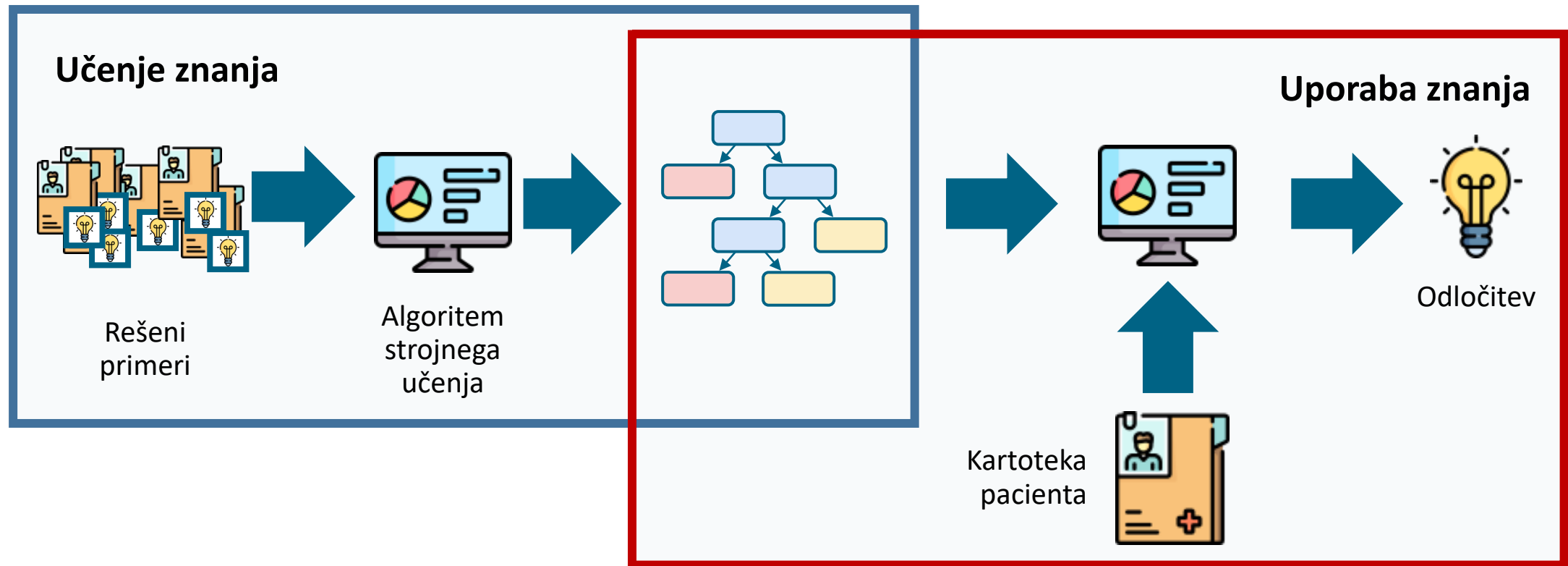
- Računalnik ne razume podatke v naši obliki. Lahko pa mu podamo informacije v za njega pravi obliki – **rešeni primeri**.



Strojno učenje



- **Inteligentni sistemi** za odločanje uporabljajo model znanja, ki so ga zgradili sami.





柯洁 KE JIE
02:58:01

ALPHAGO
02:58:55

Algoritem strojnega učenja



- Je postopek, ki **zgradi model znanja v izbrani strukturi** pri tem pa glede na vhodne podatke (podane informacije) in optimizira interno metriko kakovosti.
- Je (računsko) najpotratnejši del.
 - Največkrat je to postopek, ki po korakih postopoma prilagaja model z upanjem, da bo boljši.
- Vsak algoritem ima tudi nastavitve, s katerimi podamo navodila in omejitve pri gradnji modela (npr. velikost drevesa ali število epoh pri NN).



Struktura modela znanja	Algoritem strojnega učenja	Interna metrika kakovosti, ki se optimizira
Odločitveno drevo	CART	Gini impurity (klasifikacija), MSE (regresija)
Odločitveno drevo	ID3	Information gain
Odločitveno drevo	C4.5	Gain ratio
Matematična formula (linearni model)	OLS (klasifikacija), MLE (regresija)	Logloss ali cross-entropy (klasifikacija), MSE (regresija)
Matematična formula (nevronska mreža)	Gradient descent + backpropagation	cross-entropy
Matematična formula (nevronska mreža)	Genetic algorithm	cross-entropy, accuracy, error rate, F-score
Matematična formula (SVM)	Quadratic Programming	Hinge loss (klasifikacija), Epsilon-insensitive loss (regresija)

Interna metrika kakovosti



- Služi sprotnemu izračunu kakovosti **tekom izvajanja algoritma** strojnega učenja.
 - Če se interna metrika izračuna glede na **podane rešitve** govorimo o **nadzorovanem** (angl. *supervised*) učenju.
 - Če se interna metrika izračuna glede na **feedback oz. nagrado in kazen**, govorimo o **ojačitvenem** (angl. *reinforcement*) učenju.
 - Če se interna metrika izračuna glede na **razmerja med vhodnimi podatki**, govorimo o **nenadzorovanem** (angl. *unsupervised*) učenju.
- **Interna metrika kakovosti** ni vedno primerljiva s končno metriko kakovosti, ki je pomembna nam.
 - *gini impurity* \neq *točnost*

Metrika kakovosti končnega modela



- Uporabi se primerna glede na tip rezultat (napoved) modela in glede na tip učenja.

- **Metrika kakovosti končnega modela** ni vedno primerljiva s tem kar je pomembno nam.
 - *točnost ≠ število preživelih pacientov*



Tip rezultata modela znanja	Tip učenja	Tehnika strojnega učenja	Metrike kakovosti končnega modela
Kategorična	Nadzorovano	Klasifikacija	accuracy, error rate, F-score
Številska	Nadzorovano	Regresija	MSE, MAE, RMSE, R ²
Kategorična	Nenadzorovano	Gručenje	Silhouette, Davies-Bouldin indeks, čistoča
Kategorična	Ojačitveno	Klasifikacija	Klasifikacijske in skupna nagrada, doprinos
Številska	Ojačitveno	Regresija	Regresijske in skupna nagrada, doprinos
Več številskih	Nenadzorovano	Redukcija dimenzij	Delež pojasnene variance
Več številskih	Nenadzorovano	Preslikava	Delež napake rekonstrukcije
Kategorična	Nenadzorovano	Iskanje anomalij	Čistoča

Zakaj interna \neq končna metrika?



- **Prekomerno prileganje** (angl. *over-fitting*), ko se v model shranijo vzorci, ki so preveč značilni le za učne podatke, niso pa posplošljivi na populacijo.
 - Po domače: ko se model preveč “na pamet” nauči učne podatke.
- En izmed razlog, zakaj do tega pride je, če se preveč optimizira kakovost na učnih podatkih, brez posebnih omejitev.
 - Omejitev pri večini algoritmov strojnega učenja: interna metrika ne izraža kakovost učenja. Kako pa je pri NN?
- Kako se tega rešimo?

Upravljanje umetne inteligence



- **Upravljanje umetne inteligence** (angl. *AI alignment*) so pristopi, da sistemi, ki uporabljajo modele znanja naučenih s strojnim učenjem, zanesljivo izvajajo naloge, ki so koristne za ljudi, brez nenamernih negativnih posledic.
 - Je ključnega pomena za varno uporabo prihodnjih, bolj avtonomnih sistemov umetne inteligence.
- Zahteva reševanje **tehničnih izzivov** (kot sta robustnost in posploševanje) in **filozofskih izzivov** (kot je opredelitev človeških vrednot).