

Strojno učenje

1. Uvod u strojno učenje

prof. dr. sc. Bojana Dalbelo Bašić
doc. dr. sc. Jan Šnajder

Sveučilište u Zagrebu
Fakultet elektrotehnike i računarstva

Ak. god. 2015./2016.

Data Mining Cup 2010

Using the existing characteristics of a customer's initial order, such as order quantity per type of goods, title and delivery weight, a **decision must be made** on whether to send a voucher worth EUR 5.00. The customers who receive a voucher should be those who would not have decided to re-order by themselves.

IEEE ICDM 2010 DM Competition

Modeling the process of traffic jams formation during morning peak in the presence of roadworks, based on initial information about jams broadcast by radio stations. Input data contain identifiers of road segments closed due to roadworks, accompanied by a sequence of segments where the first jams occurred. The algorithm should **predict** a sequence of segments where next jams will occur in the nearest future.

ACM KDD Cup 1999

Learn a predictive model (i.e. a classifier) capable of **distinguishing between** legitimate and illegitimate connections in a computer network.

ACM KDD Cup 2000

Given a set of page views, will the visitor view another page on the site or will the visitor leave? Given a set of page views, **characterize** killer pages, i.e., pages after which users leave the site. Given a set of page views, **characterize** which product brand a visitor will view in the remainder of the session.

Machine Learning Challenges: <http://clopinet.com/challenges/>

Plan za prvi tjedan

- 1 Što je strojno učenje?
- 2 Literatura i internetski resursi
- 3 Pregled postupaka

Nakon predavanja, moći ćete...

- (1) definirati “strojno učenje”
- (2) opisati u kojim nam situacijama treba strojno učenje
- (3) razlikovati klasični UI-pristup rješavanju problema od strojnog učenja
- (4) nabrojati s kojim je područjima strojno učenje povezano
- (5) opisati tri osnovna pristupa strojnom učenju

- 1 Što je strojno učenje?
- 2 Literatura i internetski resursi
- 3 Pregled postupaka

Što je strojno učenje?

Strojno učenje (Alpaydin 2009)

Strojno učenje jest programiranje računala na način da **optimiziraju** neki **kriterij uspješnosti** temeljem **podatkovnih primjera** ili prethodnog iskustva.

- Raspolažemo **modelom** koji je definiran do na neke **parametre**
- Učenje: **optimizacija** parametara modela temeljem podataka

Zašto strojno učenje?

Barem tri razloga:

- 1 **Složeni problemi** – ne postoji ljudsko znanje o procesu ili ljudi ne mogu dati objašnjenje o procesu (npr. raspoznavanje govora)
 - problemi koje nije moguće riješiti na klasičan algoritamski način (*UI-potpuni problemi*)
- 2 **Ogromne količine podataka** – ima li znanja u njima?
 - otkrivanje znanja u skupovima podataka (engl. *data mining*)
- 3 **Sustavi koji se dinamički mijenjaju** – potrebna prilagodba (npr. prilagodba korisničkih sučelja)

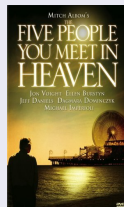
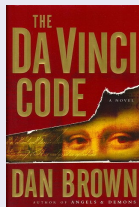
NB: To ne znači da sve probleme treba rješavati strojnim učenjem (npr. program za obračun plaća)

Od podataka do znanja

- Učenje općenitih modela iz podataka: od podataka do znanja
- **Podataka** ima u izobilju (web, tekst, eksperimentalni podatci, skladišta podataka, *deep web*, *logovi*)
- **Znanje** je skupo i potrebno
- Cilj: izgradnja modela koji je dobra i korisna **aproksimacija** podataka

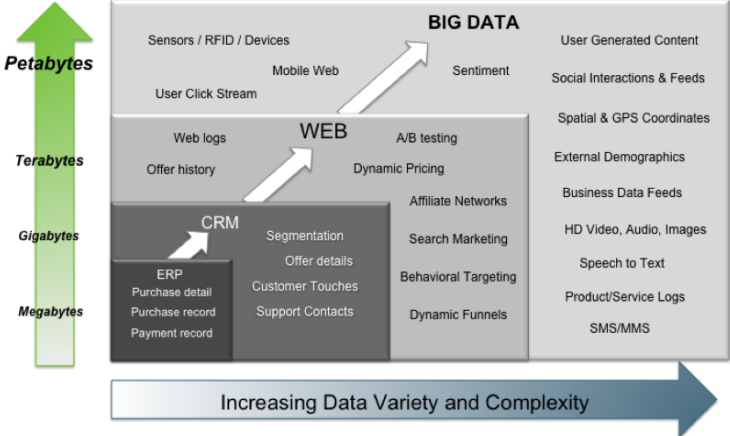
Primjer: Korisničke transakcije mogu objasniti ponašanje korisnika

*People who bought
"The da Vinci Code"
also bought
"The Five People You Meet in Heaven"*



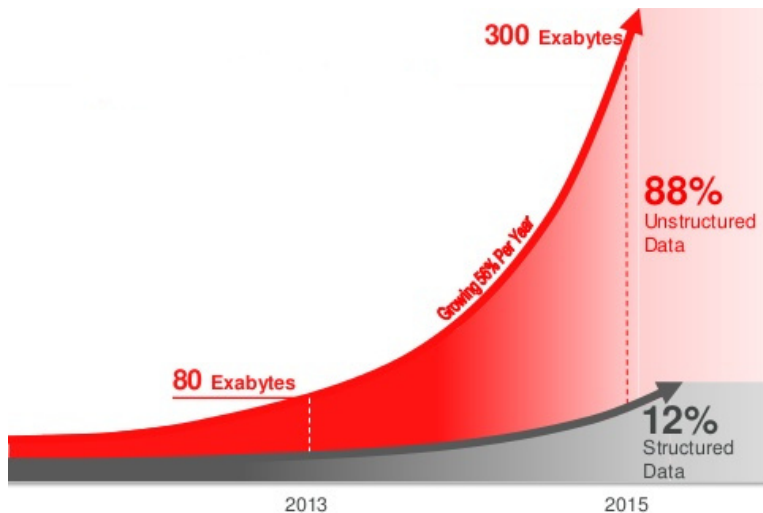
Big Data

Big Data = Transactions + Interactions + Observations

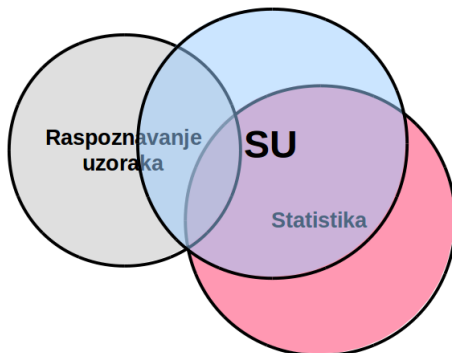


Source: Contents of above graphic created in partnership with Teradata, Inc.

(Ne)strukturirani podatci



- **Računarska znanost (umjetna inteligencija)**
- **Statistika i vjerojatnost**
- Optimizacija i upravljanje
- Numerička matematika, linearna algebra
- Teorija informacije (mjere entropije, optimalno kodiranje...)
- Filozofija (Occamova britva – najjednostavnija hipoteza je najbolja)
- Psihologija, neurobiologija, kognitivna znanost



- Intelligentni sustav treba se **prilagođavati okolini** – imati sposobnost učenja. Ako može učiti, onda može planirati ponašanje u novim situacijama
- Strojno učenje je okosnica Umjetne inteligencije
 - Robotika
 - Robotski vid
 - Raspoznavanje govora
 - Raspoznavanje uzoraka
 - Obrada prirodnog jezika: parsiranje, razrješavanje višeznačnosti, označavanje vrste riječi, ...
 - Pretraživanje informacija: rangiranje, *query log mining*
 - Umjetne neuronske mreže
 - ...



- Temeljni pojmovi u strojnom učenju: **generalizacija** i **učenje**
- Cilj: Za zadani uzorak ograničene veličine, pronaći opće pravilo koje objašnjava podatke
- **Statistika**: zaključivanje na temelju uzorka
 - generalizacija (engl. *generalization*)
→ **zaključivanje** (engl. *inference*)
 - učenje (engl. *learning*)
→ **procjena** (engl. *estimation*)



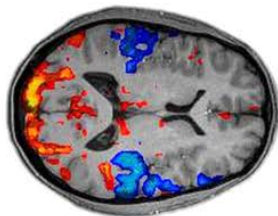
Kognitivno vjerodostojni modeli

Razumijevanje algoritama strojnog učenja \Leftrightarrow Razumijevanje ljudske sposobnosti (ili ograničenja) učenja

Thought Reading Experiment:

<http://www.cs.cmu.edu/afs/cs/project/theo-73/www/index.html>

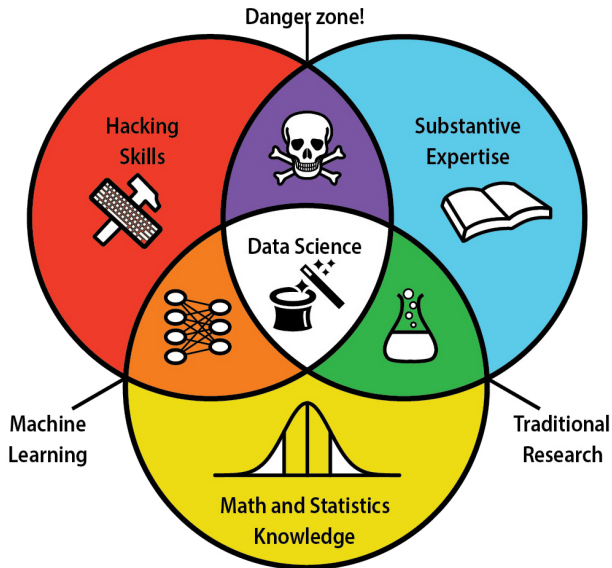
- **Funkcijska magnetna rezonancija (fMRI)**
 - Bilježi protok krvi kroz mozak: aktivna područja mozga koriste više kisika.
 - Oslanjanje na činjenicu da molekule u krvnim stanicama reagiraju u magnetskom polju u ovisnosti o količini kisika



Dubinska analiza podataka (engl. *data mining*) ili otkrivanje znanja u skupovima podataka (engl. *knowledge discovery in datasets*) – primjena strojnog učenja na velike baze podataka

- Trgovina: analiza potrošačke košarica, CRM
- Financije: Određivanje kreditne sposobnosti, detekcija zlouporaba kartica
- Proizvodnja: optimizacija, *troubleshooting*
- Medicina: postavljanje dijagnoza
- Telekomunikacije: optimizacija usluga
- Bioinformatika: analiza izražajnosti gena, poravnavanje
- Text mining: klasifikacija teksta, ekstrakcija informacija
- Računalni vid: prepoznavanje lica, praćenje vozila
- ...





The 25 Hottest Skills of 2014 on LinkedIn

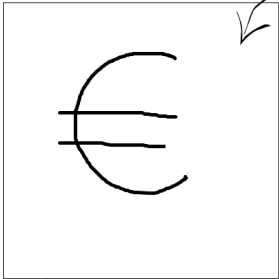
- 1 Statistical Analysis and Data Mining
- 2 Middleware and Integration Software
- 3 Storage Systems and Management
- 4 Network and Information Security
- 5 SEO/SEM Marketing
- 6 Business Intelligence
- 7 Mobile Development
- 8 Web Architecture and Development Framework
- 9 Algorithm Design
- 10 Perl/Python/Ruby
- 11 Data Engineering and Data Warehousing
- 12 Marketing Campaign Management
- 13 Mac, Linux and Unix Systems
- 14 User Interface Design
- 15 Recruiting

Jedan primjer: raspoznavanje simbola L^AT_EX-a

Detexify² - LaTeX symbol classifier

classify symbols blog

Draw here!



Did this help? Hosting Detexify costs money and if it helps you may consider helping to pay the hosting bill.

pledge.com
\$727.41 Raised!

Score: 0.0521765839556664
`\usepackage{ marvosym }`
`\EURtm`
textmode

Score: 0.0521765839556664
`\usepackage{ marvosym }`
`\EURcr`
textmode

Score: 0.0597098127426264
`\usepackage{ marvosym }`
`\EUR`
textmode

Score: 0.0600260193436406
`\usepackage{ marvosym }`
`\EURhv`
textmode

Score: 0.0705110677528292
`\usepackage{ marvosym }`
`\EURdig`
textmode

What is this?

Anyone who works with LaTeX knows how time-consuming it can be to find a symbol in [symbols-a4.pdf](#) that you just can't memorize. Detexify is an attempt to simplify this search.

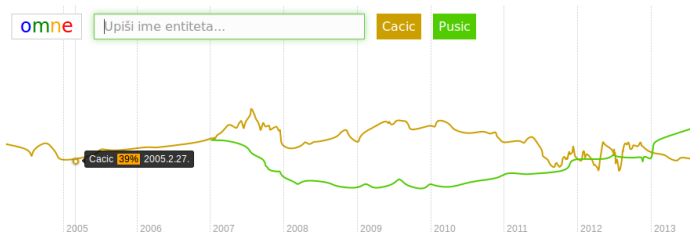
How do I use it?

Just draw the symbol you are looking for into the square area above and look what happens!

Detexify LaTeX symbol classifier (<http://detexify.kirelabs.org/>)

Još jedan primjer: analiza sentimenta

takelab.fer.hr/slavomjer



Pusic +3 -2 2014.1.11.

Pazi kakav jebeni egocentrik moras biti da nakupinu par opcinica na sjeveru ove nesretne VUKOJEBINE nayoves ni manje ni vise nego SJEVEROZAPADNI SAVEZ. Na politickoj sceni nema odumije grupe ljudi od HNSa Zelina im rascijep propast i nestajanje. Cak i u toj odumoj grupi Vesna Pusic se uspjeva istaknuti kao najgora od najgoreg. To je totalni uspjeh. Zamislite osobu pokraj koje vam Radimir Cacic izgleda kao pozitivac i cak navijate za njega. SAMO U HRVATSKOJ! Da sjeverozapadni savez. Ja cu svoj neboder nazvati STAR WARS DEATH STAR

Cacic -1 2013.12.12.

Steta sta Cacic nije prezimio s Mariborom, kakvo bi tek onda sprdanje bilo

Pusic -1 2013.11.21.

nazalost, ova vlada je potpuno nedemokratska i antiliberalna, bez obzira sto se deklarativno vole dicit tim terminima. Dapace, moglo bi se reci da naginje totalitarnoj vlasti. Sjetimo se ucjene za EU referendum usmjerene prema umirovljenicima, a sada pad Pusic kaze glasajte protiv bez obzira sto mislite o tome. Mislim da se gradani trebaju osvjestiti, jer svaki uspjeh u manipulacijama vlade im daje vjetar u leda za nove poteze protiv naroda.

Cacic +1 -1 2013.10.15.

Svi ste zaboravili da je Cacic slobodan, a repka jos uvijek igra na Maksimiru

Cacic -2 2013.9.14.

Rasa nije state of mind nego fenotipsko nasljeđe. Znacni nema nikakve razlike među nama osim toga. Pricati o međurasnim obilježjima možemo pričati samo u kontekstu kulture i to na tlu gdje neka rasa prevladava. Slazem se da Belgija i Francuska imaju gadnih problema i već sam rekla da radikalizam treba izolirati i odstraniti. Država treba raditi

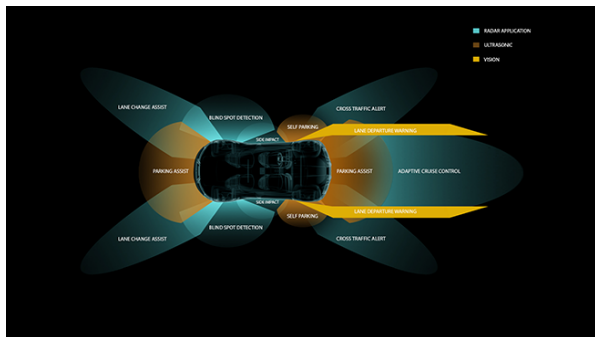
Duboko učenje: sljedeća velika stvar u SU?

- Pristup strojnom učenju temeljen na neuronskim mrežama i velikoj količini podataka
- Takav pristup je sada moguć zbog enormne količine podataka generirane od strane (internet) korisnika (*user-generated content*)
- Postiže *state-of-the-art* rezultate na širokom spektru zadataka
- Google, Baidu, Paypal i NVIDIA

Duboko učenje: autonomno vozilo (NVIDIA)

<http://blogs.nvidia.com/blog/2015/02/24/deep-learning-drive/>

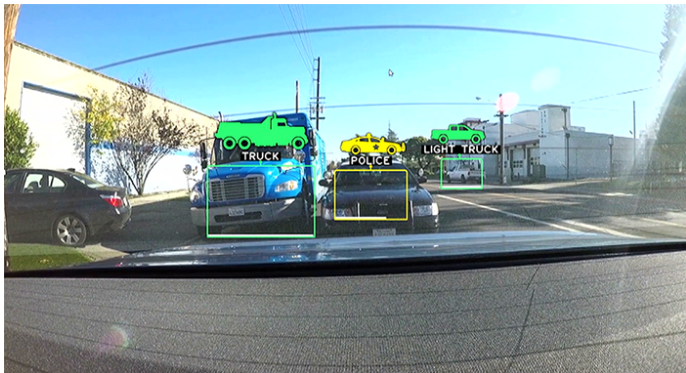
- Razni senzori skupljaju veliku količinu podataka koje DU koristi za razne izazove unutar automatske vožnje
- Prepoznavanje pješaka, automatsko parkiranje, pomoć pri prestrojavanju itd.



Dubinsko učenje: autonomno vozilo (NVIDIA)

<http://blogs.nvidia.com/blog/2015/02/24/deep-learning-drive/>

- Klasifikator raspoznaje razne vrste vozila i prometnih znakova, čak i kada je čovjeku teško obaviti taj zadatak (kada je nevjeme i sl.)



Dubinsko učenje: prepoznavanje broja kalorija (Google)

<http://www.popsci.com/google-using-ai-count-calories-food-photos>

- Sustav na temelju fotografije jela procjenjuje koliko ono sadrži kalorija
- Zasad u povojima, ali vrlo perspektivno



System A: Student angel watch university care program

- **Što?**

Za savjetodavnu službu FER-a treba razviti sustav ranog upozoravanja od pada godine

- **Zašto?**

Povećati uspjeh studiranja i smanjiti vrijeme studiranja, time doprinijeti ekonomiji i sreći stanovništva

- **Ulaz:**

?

- **Izlaz:**

Upozorenje da će student/ica X vjerojatno pasti godinu u skoroj budućnosti ako tako nastavi

System B: Box-office oracle

- **Što?**

Za Warner Bros. Entertainment Inc. razviti sustav koji nalazi parametre produkcije koji maksimiziraju dobit

- **Zašto?**

Zato jer gledatelji zaslužuju najbolje :-)

- **Ulaz:**

?

- **Izlaz:**

Predviđena ukupna dobit filma za zadane parametre produkcije

System C: Psychic smartphone deejay

- **Što?**

Za Google Inc. razviti sustav koji odabire prikladnu muziku s obzirom na ono što korisnik trenutno radi na smartphoneu

- **Zašto?**

Radi sinestetičkog doživljaja, i zato jer Google zna bolje od vas što vam treba

- **Ulaz:**

?

- **Izlaz:**

Rang-lista tri najprikladnije pjesme za zadanu situaciju

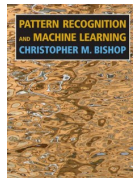
Kako?

...3 minute...

Plan za prvi tjedan

- 1 Što je strojno učenje?
- 2 Literatura i internetski resursi
- 3 Pregled postupaka

- Ethem Alpaydin: *Introduction to Machine Learning*, MIT Press, 2009.
- Christopher Bishop: *Pattern Recognition and Machine Learning*, Springer, 2007.
- Kevin P. Murphy: *Machine Learning: A Probabilistic Perspective*, MIT Press, 2012.



- Stephen Marsland: *Machine Learning: An Algorithmic Perspective*, Chapman and Hall/CRC, 2009.
- Duda, Hart, Stork: *Pattern Classification*, Wiley-Interscience, 2000.
- Mohri, Rostamizadeh, Talwalkar: *Foundations of Machine Learning*, MIT Press, 2012.
- Witten, Frank, Hall: *Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kaufmann, 2011.
- Hastie, Tibshirani, Friedman: *Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer, 2003.
- Daphne Koller: *Probabilistic Graphical Models: Principles and Techniques*, MIT Press, 2009.
- Tom Mitchell: *Machine Learning*, McGraw-Hill, 1997.

- Hastie, Tibshirani, Friedman: *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*
<http://www-stat.stanford.edu/~tibs/ElemStatLearn/>
- MacKay: *Information Theory, Inference, and Learning Algorithms*
<http://www.inference.phy.cam.ac.uk/mackay/itila/book.html>
- Sutton & Barto. *Reinforcement Learning: An Introduction*
<http://webdocs.cs.ualberta.ca/~sutton/book/ebook/>
- Rasmussen & Williams: *Gaussian Processes for Machine Learning*
<http://www.gaussianprocess.org/gpml/chapters/>
- Barber: *Bayesian Reasoning and Machine Learning*
<http://www.cs.ucl.ac.uk/staff/d.barber/brml>
- Nilsson: *Introduction to Machine Learning*
<http://ai.stanford.edu/~nilsson/mlbook.html>

- International Conference on Machine Learning (ICML)
<http://www.machinelearning.org/icml.html>
- European Conference on Machine Learning (ECML)
ECML11: <http://www.ecmlpkdd2011.org/>
- Neural Information Processing Systems (NIPS)
<http://nips.cc/Conferences/>
- Uncertainty in Artificial Intelligence (UAI)
<http://www.auai.org/>
- Computational Learning Theory (COLT)
<http://www.learningtheory.org/>
- International Joint Conference on Artificial Intelligence (IJCAI)
<http://ijcai.org/>
- International Conference on Neural Networks (Europe)
ICANN11: <http://www.cis.hut.fi/icann2011/>

- Journal of Machine Learning Research (www.jmlr.org)
- Machine Learning (www.springer.com/computer/ai/journal/10994)
- Neural Computation
- Neural Networks
- IEEE Transactions on Neural Networks
- IEEE Transactions on Pattern Analysis and Machine Intelligence
- Annals of Statistics
- Journal of the American Statistical Association
- ...

- MetaOptimize QA
Strojno učenje, NLP, AI, IR, vizualizacija i analiza podataka
<http://metaoptimize.com/>
- CrossValidated QA
Statistika, dubinska analiza i vizualizacija podataka
<http://stats.stackexchange.com/>
- KD nuggets
<http://www.kdnuggets.com/>
- Data Mining Cup
<http://www.data-mining-cup.de/en/>
- Machine Learning Summer School (MLSS)
<http://www.mlss.cc/>
- Open Source Machine Learning (MLOSS)
<http://mloss.org>

- Videlectures.net
videlectures.net/Top/Computer_Science/Machine_Learning/
- Andrew Ng (Stanford): Machine Learning lectures
academicearth.org/courses/machine-learning
- Coursera courses (<https://www.coursera.org/>)
 - Machine Learning:
<https://www.coursera.org/course/ml>
 - Probabilistic Graphical Models:
<https://www.coursera.org/course/pgm>

- Weka (GPL)

www.cs.waikato.ac.nz/ml/weka

- Rapid Minner (AGPL)

rapid-i.com

- Orange (GPL)

www.ailab.si/orange

- R (GPL)

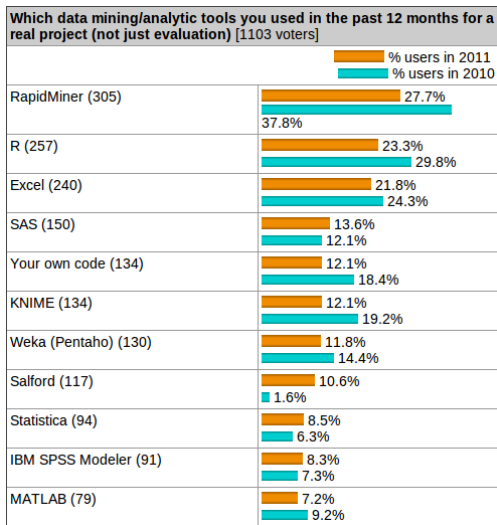
www.r-project.org

- Matlab

www.mathworks.com/products/matlab



Programski alati – popularnost














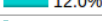

Izvor: KDnuggets polls (<http://www.kdnuggets.com/2011/05/tools-used-analytics-data-mining.html>)

SciPy ecosystem



What languages you used for data mining / data analysis?

What programming languages you used for data mining / data analysis in the past 12 months? [570 voters]

R (257)	 45%
SQL (184)	 32%
Python (140)	 25%
Java (139)	 24%
SAS (121)	 21%
MATLAB (83)	 15%
C/C++ (73)	 13%
Unix shell/awk/gawk/sed (59)	 10%
Perl (45)	 7.9%
Hadoop/Pig/Hive (35)	 6.1%
Lisp (4)	 0.7%
Other (70)	 12.0%
None (7)	 1.2%

Izvor: KDnuggets polls (<http://www.kdnuggets.com/2011/05/tools-used-analytics-data-mining.html>)

- UCI Repository
<http://www.ics.uci.edu/~mlearn/MLRepository.html>
- UCI KDD Archive
<http://kdd.ics.uci.edu/summary.data.application.html>
- Statlib
<http://lib.stat.cmu.edu/>
- Weka datasets
http://www.cs.waikato.ac.nz/ml/weka/index_datasets.html
- Delve
<http://www.cs.utoronto.ca/~delve/>

SU2010



- Web-stranica konferencije:
<http://www.fer.unizg.hr/predmet/su/su2010>
- Zbornik radova:
http://www.fer.hr/_download/repository/su2010-zbornik.pdf

Plan za prvi tjedan

- 1 Što je strojno učenje?
- 2 Literatura i internetski resursi
- 3 Pregled postupaka

Vrste strojnog učenja

(1) Nadzirano učenje (engl. *supervised learning*)

Podatci su u obliku (*ulaz, izlaz*)= (x, y) , treba naći preslikavanje $\hat{y} = f(x)$

- Ako je y je diskretna/nebrojčana vrijednost: **klasifikacija**
- Ako je y kontinuirana/brojčana vrijednost: **regresija**

(2) Nenadzirano učenje (engl. *unsupervised learning*)

Dani su podaci bez ciljane vrijednosti, treba naći pravilnost u podacima

- **grupiranje** (engl. *clustering*)
- **procjena gustoće** (engl. *density estimation*)
- **smanjenje dimenzionalnosti** (engl. *dimensionality reduction*)

(3) Podržano/ojačano učenje (engl. *reinforcement learning*)

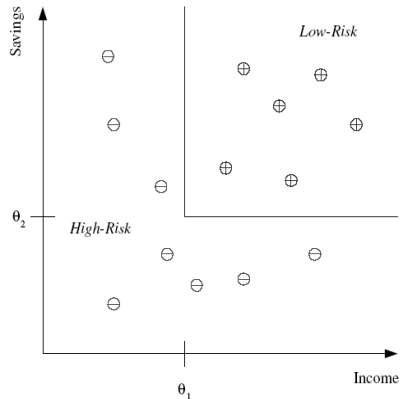
Učenje optimalne strategije na temelju pokušaja s odgođenom nagradom

(1) Nadzirano učenje

- **Predviđanje:** na temelju ulaznih vrijednosti predvidjeti buduće
- **Ekstrakcija znanja:** učenje lako tumačivih modela
- **Sažimanje:** model koji koncizno objašnjava podatke
- **Otkrivanje ekstremnih vrijednosti:** iznimke koje nisu pokrivena modelom (npr. zlouporaba)
- **Upravljanje:** upravljački ulazi dobiveni kao izlaz regresije

Primjer klasifikacije: Analiza kreditne sposobnosti

- Klasifikacija u svrhu **predviđanja**
- Cilj: Razlikovanje između grupa klijenata niskog rizika i visokog rizika na temelju podataka o njihovom prihodu i uštedevini
- Diskriminacijska funkcija:
IF $\text{prihod} > \theta_1$ **AND**
 uštedevina $> \theta_2$
THEN nizak rizik
ELSE visok rizik

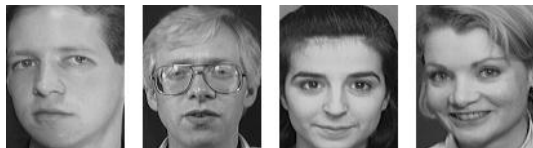


Primjer klasifikacije: raspoznavanje lica

- Cilj: prepoznati lice osobe unatoč promjenama u pozici, osvjetljenju, frizuri, šminki te okluzijama (naočale, brada)
- Skup podataka za učenje:



- Budući podatci:



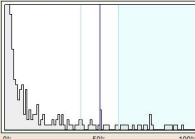
- Baze lica: <http://www.face-rec.org/databases/>

Primjer klasifikacije: kategorizacija novinskih članaka

Filter: One category view

Showing category 'POLITIKA|IZBORI|Parlamentarni izbori' from collection 'Collection0002'

Label distribution



0% 50% 100%

Show documents

Negative Uncertain Positive

Labeled: 0 0 5

Unlabeled: 4954 18 15

Manual label	Auto output	Auto rule	Id
Yes	0.931	0	97273
Yes	0.908	0	96049
Yes	0.899	0	98242
Yes	0.874	0	94597
Yes	0.869	0	97914
Yes	0.855	0	95792
No	0.844	0	93620
	0.778	0	97730
	0.769	0	97458
	0.767	0	94764
	0.766	0	96628
	0.750	0	97631
	0.733	0	96532
	0.715	0	96476
	0.704	0	96606
	0.676	0	97643
	0.646	0	93932
	0.636	0	94729
	0.627	0	97640

Article

Publisher: Glas Slavonije Date: 20.06.2005.

Vlada namjerava zabraniti koalicije prije izbora? CELNICI VLADE RAZMATRAJU PROMJENE IZBORNIH PRAVILA

Vlada namjerava zabraniti koalicije prije izbora? Neslužbeno, HDZ za sljedeće izbore namjerava predložiti ili zabranu predizborno koaliciranja, ili povišenje praga na osam posto za koaliciju dvije, odnosno 11 posto za koaliciju tri ili više stranaka ZAGREB - Vlada posljednjih dana intenzivno razmatra mogućnosti izmjena izbornog zakona kojima bi se omogućile manipulacije kakve su se pojavile nakon nedavnih lokalnih izbora, potvrdio nam je visoki vladin dužnosnik. Kako se neslužbeno doznaje, HDZ za sljedeće parlamentarne i lokalne izbore namjerava predložiti ili zabranu predizborno koaliciranja, ili povišenje praga na osam posto za koaliciju dvije, odnosno 11 posto za koaliciju tri ili više stranaka. Zabrana koaliciranja prije izbora značila bi prihvatanje njemačkog modela, a povišenje izbornog praga povratka na izborna pravila po kojima su se u Hrvatskoj izbori održavali u devedesetima. Neslužbeno se može čuti da u Vladu štuju signali iz Europske komisije da konačno zakonski onemogući cirkuse s kupovanjima vijećnika i zastupnika, kojih smo bili svjedoci posljednjih tjedana. Vlada vjerojatno ipak neće brzati s izmjenama izbornih zakona zbog toga jer do redovitih parlamentarnih izbora ima dvije i pol, a do lokalnih čak četiri godine. No, HDZ-u je vjerojatno u interesu da se sljedeći parlamentarni izbori, ako oni budu i prijevremeni, održe po novim pravilima. Visoki vladin dužnosnik otkriva nam da Banski dvori posljednjih dana dobivaju mnogo zahtjeva s terena, i to ne samo od HDZ-a nego i ostalih stranaka, da se utvrde nova pravila i onemogući lakdija na budućim izborima.

TEŠKO IZVEDIVA ZABRANA

Promjena izbornih pravila ponajprije bi pogodovala takozvanim velikim strankama, primjerice HDZ-u i SDP-u, jer bi pokupili bardo glasova koji se rasprše na male stranke, umjesto da nakon izbora pokušavaju privoljeti na suradnju njihove, često nepouzdane, zastupnike i vijećnike kako bi osigurali većinu.

Zastupnik SDP-a Mato Arlović kaže kako bi zabrana predizborno koaliciranja bila teško izvediva, ali i da je u svojoj biti suprotna ustavnim odredbama koje jamče višestranački sustav. Stoga Arlović smatra da ne treba zabranjivati predizborne koalicije, ali da svakako treba otežati njihovo formiranje. - Trebalo bi čiti na proporcionalni sustav, ali onaj koji bi omogućio građanima da mogu glasovati i za ljude i za njihov redoslijed na listi, a ne samo za stranačke liste. Da o tome iko će na kraju biti u Saboru odlučuju birači, a ne stranačka vodstva, kaže Arlović. Što se izbornih biračkih listica odnosi na 11 posto, Arlović kaže da to ne

URL: -
Extern Id: -

Relative impacts

Feature	Impact
izbor	2.091
stranka	1.251
predizboran	1.033
izboran	0.904
parlamenta.	0.579
koalicija	0.515
izboriti	0.432
zastupnik	0.275
vlada	0.243
hdz	0.209
birac	0.155
jednica	0.120
glasovati	0.108
glas	0.105
namjeravati	0.097
izici	0.097
pravilo	0.092
onemogućiti	0.072
praviti	0.068
politički	0.058
manipulacija	0.054
sdp	0.050
taština	0.048
jamčiti	0.040
promjena	0.035
sabor	0.034
osigurati	0.033
vodstvo	0.029
čelnik	0.028
zabrana	0.028
signal	0.025
ustavan	0.022
kazati	0.019
odlučivati	0.019
kupovanje	0.016
kupovan	0.016
čuti	0.015
sljedeći	0.014
stizati	0.012
matr	0.011

Save Cancel

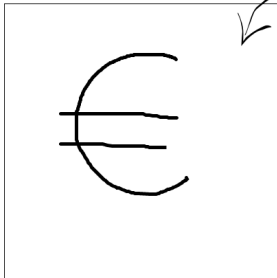
TakeLab KTN indexer (<http://takelab.fer.hr/en/products/63-ktn-indexing-system>)

Detexify² - LaTeX symbol classifier

classify

symbols

blog



clear

What is this?

Anyone who works with LaTeX knows how time-consuming it can be to find a symbol in [symbols-a4.pdf](#) that you just can't memorize. Detexify is an attempt to simplify this search.

How do I use it?

Just draw the symbol you are looking for into the square area above and look what happens!

Draw here!

Did this help?

Hosting Detexify costs money and if it helps you may consider helping to pay the hosting bill.



Score: 0.0521765839556664

```
\usepackage{ marvosym }
```

```
\EURtm
```

```
textmode
```



Score: 0.0521765839556664

```
\usepackage{ marvosym }
```

```
\EURcr
```

```
textmode
```



Score: 0.0597098127426264

```
\usepackage{ marvosym }
```

```
\EUR
```

```
textmode
```



Score: 0.0600260193436406

```
\usepackage{ marvosym }
```

```
\EURhv
```

```
textmode
```



Score: 0.0705110677528292

```
\usepackage{ marvosym }
```

```
\EURdig
```

```
textmode
```

Detexify LaTeX symbol classifier (<http://detexify.kirelabs.org/>)

Primjer klasifikacije: raspoznavanje novčanica

Klasifikacija novčanica

5 kn | 10 kn | 20 kn | 50 kn

Istrosenost 88%

Poderanost 52%

Izguzvanost 90%

Necistoce 10%

Isaranost 0%

Generiraj uzorak

Lice Rotirano

Nalije Setep

Mreza 990x3x4

uzorak 45x22

ulazni sloj

skiveni sloj

izlazni sloj

Propagiraj faktori_1_17.annf

0.01 5 kn

0.01 10 kn

0.98 20 kn

0.00 50 kn

Neuronska mreža i generator uzoraka (<http://www.zemris.fer.hr/predmeti/su/seminari/>)

Primjer klasifikacije: predviđanje ishoda izbora



Election Analytics (<http://electionanalytics.cs.illinois.edu/index.html>)

Primjer klasifikacije: predviđanje suicidalnosti



Predicting Suicide Intent (PSI) Proposers' Day Workshop

Solicitation Number: DARPA-SN-12-56
Agency: Other Defense Agencies
Office: Defense Advanced Research Projects Agency
Location: Contracts Management Office

Notice Details

Packages

Interested Vendors List

Print Link

Original Synopsis

Sep 26, 2012
3:06 pm

[Return To Opportunities List](#)

[Watch This Opportunity](#)

[Add Me To Interested Vendors](#)

Solicitation Number: DARPA-SN-12-56
Notice Type: Special Notice

Synopsis:

Added: Sep 26, 2012 3:06 pm

The Defense Sciences Office (DSO) of the Defense Advanced Research Projects Agency (DARPA) is hosting an Understanding and Preventing Suicide Workshop to discuss scientific, mathematical and computational models that predict when a person is likely to commit suicide so as to provide opportunity for clinicians to institute appropriate medical interventions. See attached DARPA-SN-12-56.

ALL FILES

[Package](#)
Sep 26, 2012

[DARPA-SN-12-56.pdf](#)

GENERAL INFORMATION

Notice Type: Special Notice
Posted Date: September 26, 2012
Response Date: Oct 12, 2012 11:59 pm Eastern
Archiving Policy: Automatic, on specified date
Archive Date: October 17, 2012
Original Set Aside: N/A

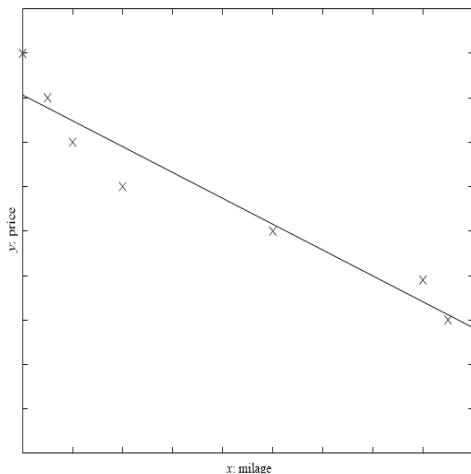
(<http://www.livescience.com/23552-military-minority-report-suicide.html>)

- Klasifikacija novinskih dokumenata u rubrike
- Detekcija neželjenih poruka e-pošte (engl. *spam detection*)
- Predviđanje kretanja dionica
- Određivanje smisla višeznačne riječi
- Raspoznavanje dlanova u svrhu autentikacije
- Automatsko dodjeljivanje ključnih riječi nekom dokumentu
- Medicinska dijagnostika (od simptoma do dijagnoze ili obrnuto)
- Prepoznavanje vrste plesa na temelju ritma
- Predviđanje ishoda nogometnih utakmica
- Klasificiranje mentalnog zdravlja autora teksta
- Predviđanje ocjene filma na temelju ocjena gledatelja

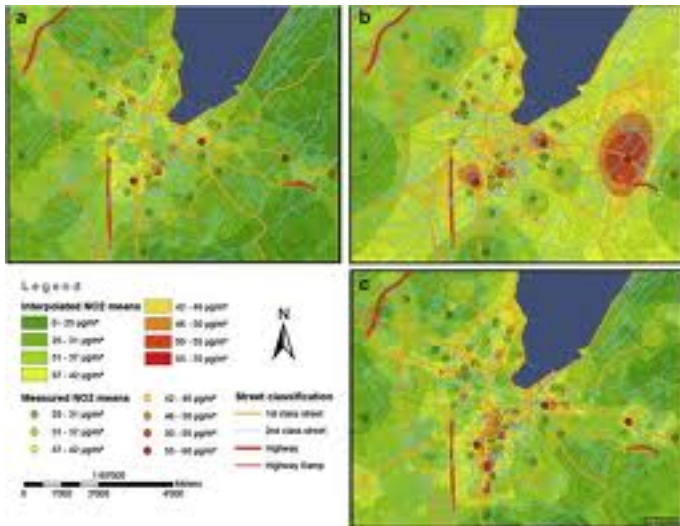
Regresija – primjer

Cilj: Predviđanje cijene rabljenih automobila

- Atributi:
 - x – prijeđeni km
 - y – cijena
- Aproksimacija cijene:
 $\hat{y} = h(x)$
- Model:
 $h(x) = w_1x + w_0$
 w_0, w_1 – parametri modela



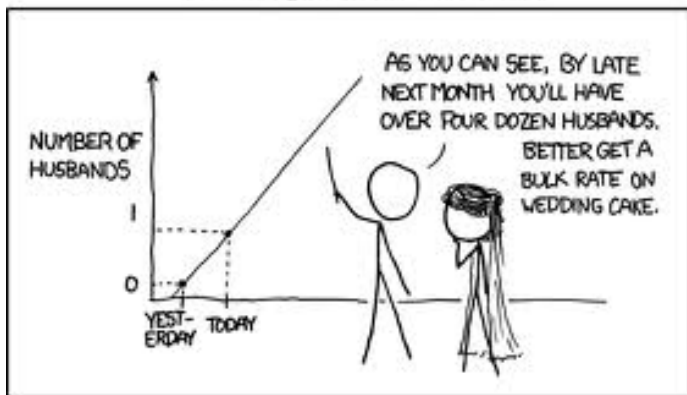
Primjer regresije: predviđanje razine emisije NO₂



Exposure models for traffic related NO₂ (<http://www.sciencedirect.com/science/article/pii/S1352231011009629>)

Model: dobra i korisna aproksimacija!

MY HOBBY: EXTRAPOLATING



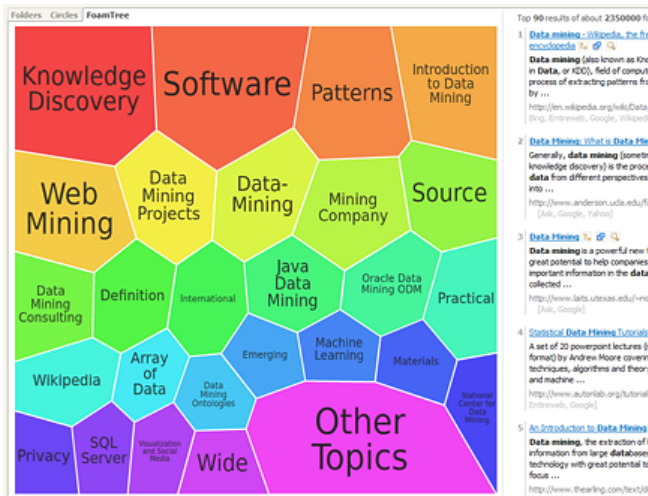
(2) Nenadzirano učenje

- Dani su podaci, bez ciljne vrijednosti – **neoznačeni podaci** (engl. *unlabeled data*)
- Cilj nenadziranog učenja jest naći pravilnosti u podacima
- Tipične primjene:
 - Eksplorativna dubinska analiza podataka
 - Marketing: segmentacija korisnika
 - Biologija: grupiranje organizama prema njihovim značajkama
 - Text mining: grupiranje sličnih dokumenata
 - Pretraživanje informacija: grupiranje sličnih rezultata
 - Bioinformatika: grupiranje DNA-mikropolja
 - Obrada slike: sažimanje slike grupiranjem istobojnih piksela

Grupiranje (engl. *clustering*)

Razvrstavanje podataka u grupe tako da slični podatci budu u istoj grupi, a različiti podatci u različitim grupama.

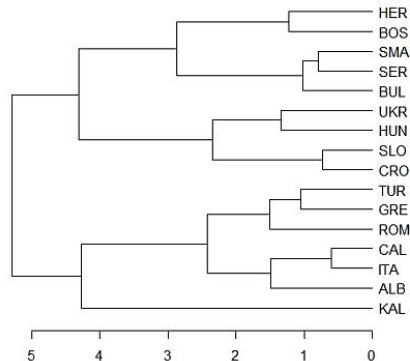
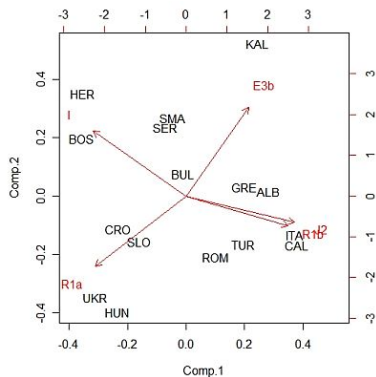
Primjer: grupiranje rezultata pretraživanja



Carrot² – Open Source Search Results Clustering Engine (<http://project.carrot2.org/>)

Primjer: grupiranje haploskupina (evolucijska biologija)

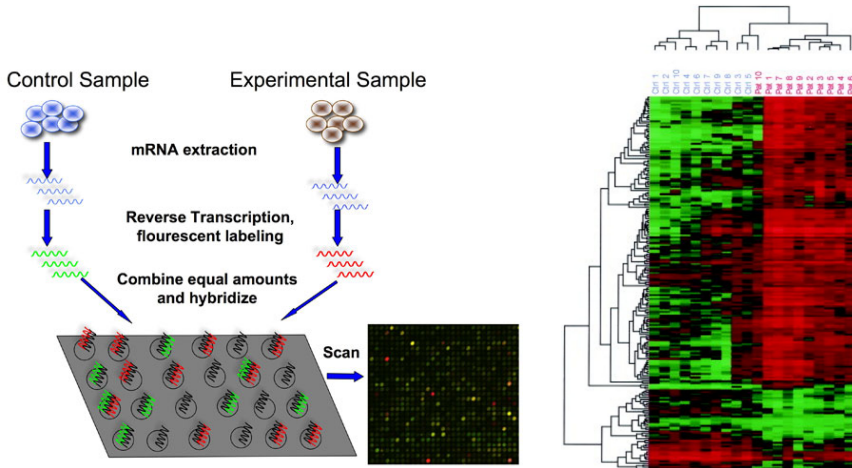
Haploskupine – nasljedno, polovično genetičko obilježje, korisno za analizu genetičkog podrijetla populacija



Dienekes' Anthropology Blog, <http://dienekes.blogspot.com/2005/08/haplogroup-frequency-correlations-in.html>

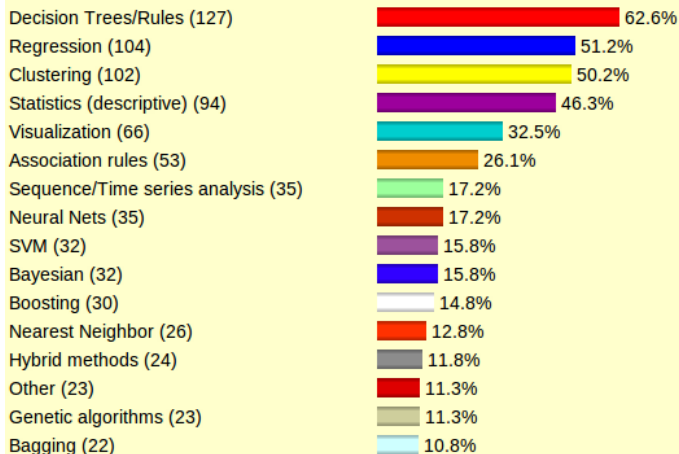
Primjer: grupiranje DNA-mikropolja (bioinformatika)

Cilj: grupiranje gena sa sličnom izražajnošću
(slična izražajnost – slična funkcionalnost)



Algoritmi – popularnost (dubinska analiza podataka)

Data mining/analytic methods you used frequently in the past 12 months: [203 voters]



Izvor: KDnuggets polls 2007 (http://www.kdnuggets.com/polls/2007/data_mining_methods.htm)

(3) Podržano učenje

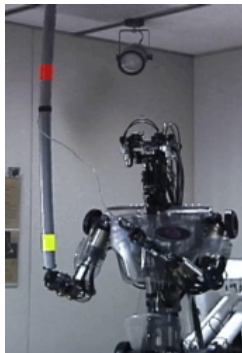
- Učenje strategije na temelju serije izlaza
 - Nema nadziranog učenja – samo odgođena nagrada
 - Problem dodjeljivanja nagrade (engl. *credit assignment problem*)
-
- Tipične primjene:
 - Igranje igara
 - Robotika i upravljanje
 - Višeagentni sustavi



Primjer: upravljanje humanoidnim robotom

Cilj: generiranje upravljačkih akcija za humanoidnog robota

- problem: 7 ili više stupnjeva slobode (npr. zglobovi ruke)
- prostor stanja ima 21 ili više dimenzija (po jednu za položaj, brzinu i akceleraciju svakog zgloba)



TU Darmstadt: Intelligent Autonomous Systems (<http://www.robot-learning.de/Research/ReinforcementLearning>)

- **Strojno učenje** bavi se rješavanjem problema koje je teško riješiti klasičnim pristupima
- Podataka ima u izobilju – izazov je transformirati podatke u u **znanje** tako da možemo **zaključivati i predviđati**
- Strojno učenje povezano je sa statistikom, kognitivnom znanosti, računarskom znanosti, umjetnom inteligencijom, raspoznavanjem uzoraka, ...
- Učenje se svodi na **optimizaciju parametara modela** na temelju **podataka**
- Tri osnovna pristupa: **nadzirano učenje** (klasifikacija i regresija), **nenadzirano učenje** i **podržano učenje**
- Organizacijski: tjedne domaće zadaće (ne boduju se), šest labosa u SciPy-u (pet ih se boduje), ispit, Kaggle za ambiciozne



Sljedeća tema: Nadzirano učenje

Sada možete....

- (1) definirati “strojno učenje”
- (2) opisati u kojim nam situacijama treba strojno učenje
- (3) razlikovati klasičan UI-pristupa rješavanju problema od strojnog učenja
- (4) nabrojati s kojim je područjima strojno učenje povezano
- (5) opisati tri osnovna pristupa strojnom učenju