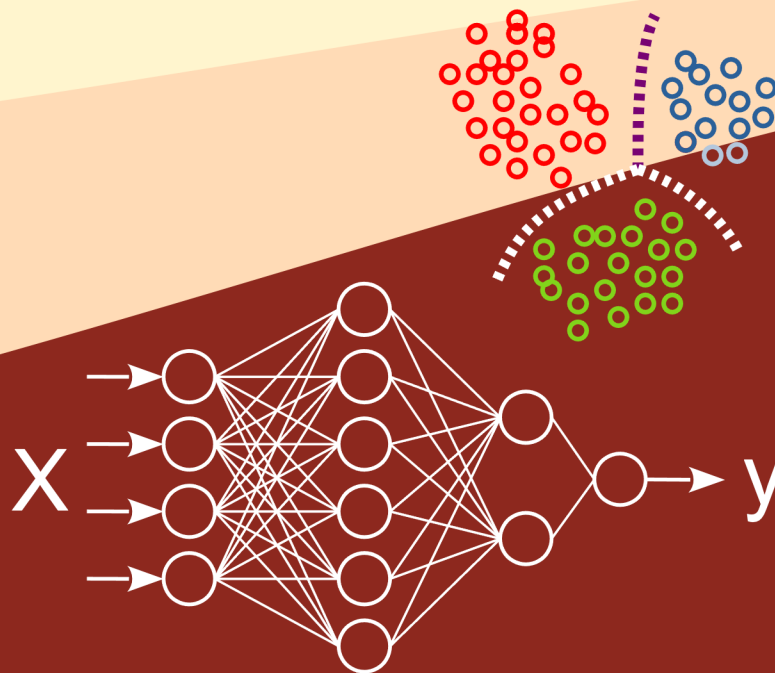
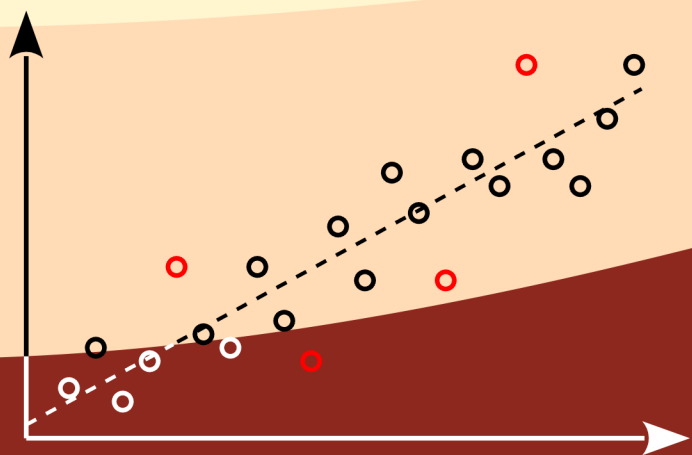


Arhitektura i Razvoj Inteligentnih Sustava

Tjedan 1: Uvodno predavanje



Creative Commons



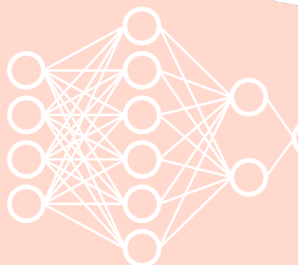
- slobodno smijete:

- dijeliti — umnožavati, distribuirati i javnosti priopćavati djelo
- prerađivati djelo



- pod sljedećim uvjetima:

- imenovanje: morate priznati i označiti autorstvo djela na način kako je specificirao autor ili davatelj licence (ali ne način koji bi sugerirao da Vi ili Vaše korištenje njegova djela imate njegovu izravnu podršku).
- nekomercijalno: ovo djelo ne smijete koristiti u komercijalne svrhe.
- dijeli pod istim uvjetima: ako ovo djelo izmijenite, preoblikujete ili stvarate koristeći ga, prerađu možete distribuirati samo pod licencom koja je ista ili slična ovoj.



U slučaju daljnjeg korištenja ili distribuiranja morate drugima jasno dati do znanja licencne uvjete ovog djela.

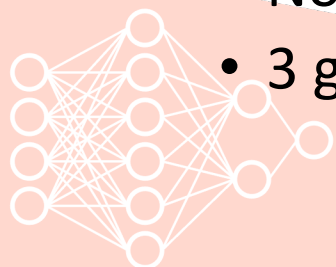
Od svakog od gornjih uvjeta moguće je odstupiti, ako dobijete dopuštenje nositelja autorskog prava.

Ništa u ovoj licenci ne narušava ili ograničava autorova moralna prava.

Tekst licence preuzet je s <http://creativecommons.org/>

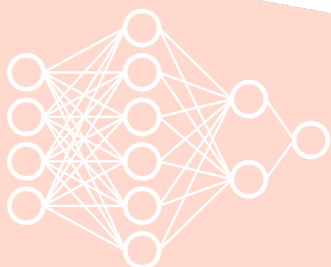
O predmetu

- Izborni diplomski predmet
- 5 ECTS bodova
- 30 sati predavanja
- 15 sati laboratorijskih vježbi – svih 5 je fizičkih vježbi
 - Bodovi se dobiju za prisutnost na laboratorijskoj vježbi – 20% bodova
- Bodovi se dobiju i za prisutnost na predavanjima – 10% bodova
- Predmet se polaže kroz projekt – 70% bodova
 - Projekt će biti izazovan i sadržat će sve elemente MLOps-a
 - Nosi 70% bodova i biti će pregledan od strane nastavnika na kraju semestra
 - 3 grupe po 7 studenata



O predmetu

- Predmet je sadrži teme praktičnog strojnog učenja (*ML engineering*) i zamišljen je kao praktičan predmet
- Ideja je približiti vas poslovnim i tehničkim detaljima uvođenja sustava temeljenih na strojnom učenju – inteligentnih sustava
- Tijekom predmeta obraditi će se teme:
 - Arhitektura inteligentnih sustava
 - Upotreba inteligentnih sustava u poslovanju
 - Programiranje inteligentnih sustava i infrastrukturni detalji
 - MLOps metodologija

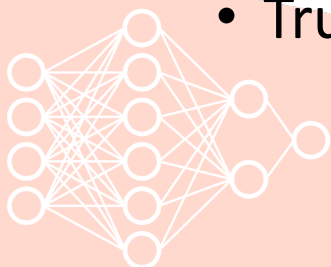


Nastavno osoblje

- Nositelj: doc.dr.sc. Mario Brčić, mario.brcic@fer.hr
 - Projekt
- Predavač: dr.sc. Dalibor Krleža, dalibor.krleza@fer.hr
 - Predavanja i laboratorijske vježbe
- Asistent: Andrija Miličević, mag.ing., andrija.milicevic@fer.hr
 - Laboratorijske vježbe
- Vanjski suradnik: Zoran Krleža, mag.ing., zoran.krleza@true-north.hr
 - RedHat certified architect
 - Kubernetes, docker, infrastruktura za projekt
 - True North d.o.o.



TRUE NORTH

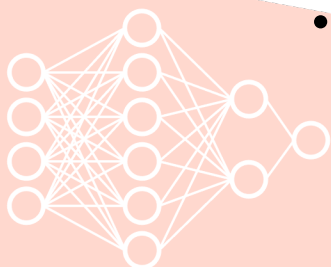


O nama i predmetu

- Naša namjera – kontekst ovog predmeta:

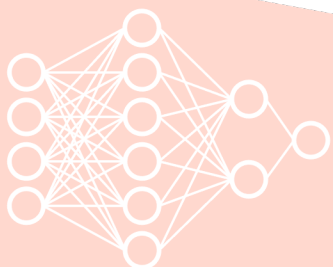
"Provesti Vas glatko kroz sadržaj predmeta i opremiti Vas znanjima koja će dobro prezentirati Vas i FER u svijetu."

- Imajte na umu da se ovaj predmet na FER-u održava po prvi puta
 - Sličnih predmeta ima po drugim US sveučilištima – pojavljuju se od 2019.
 - Od vas tražimo nove ideje za predmet
 - Što je bilo dobro i što zadržati?
 - Što se još očekivali čuti?



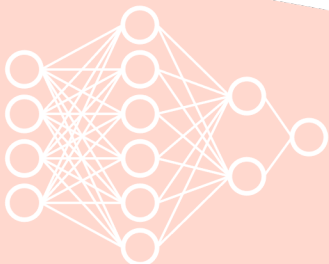
Raspored predavanja (1)

| Tjedan | Datum | Predavanja | Lab | Predavač(i) |
|--------|--------|---|---|-------------------------------|
| 1 | 2.3. | <ul style="list-style-type: none"> Uvodno predavanje Zašto i čemu inteligentni sustavi? Problematika uvođenja i zašto mnogo takvih projekata propada. | Instalacija i osposobljavanje razvojne okoline za laboratorijske vježbe | Mario Brčić Dalibor Krleža |
| 2 | 9.3. | <ul style="list-style-type: none"> Arhitektura klasičnih sustava kao i razlike prema inteligentnim sustavima Poslovni slučajevi korištenja inteligentnih sustava Klasifikacija i algoritmi Model kao osnova inteligentnog sustava | | Dalibor Krleža |
| 3 | 13.03. | | Lab1: Osnove, model, spremanje i učitavanje modela, metrika i evaluacija modela | |
| | 16.03. | <ul style="list-style-type: none"> Zahtjevi na infrastrukturu Ocjena razlika i potrebnih nadogradnji infrastrukture | | Dalibor Krleža |



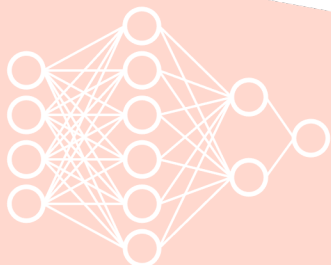
Raspored predavanja (2)

| | | | |
|---|--------|---|-------------------------------|
| 4 | 23.03. | <ul style="list-style-type: none">• Preporučivači (recommenderi) kao slučaj korištenja• Zadatak za projekt• Izvori podataka• Podatkovna integracija• Dohvat podataka iz velikih tokova podataka | Mario Brčić Dalibor Krleža |
| | | <ul style="list-style-type: none">• Postupci pripreme i čišćenja podataka• Automatizacija transformacije podataka (ETL)• Velike količine podataka i skalabilnost transformacije• Odabir algoritma strojnog učenja s obzirom na podatke, te utjecaj odabira algoritma na postupak pripreme i čišćenja | Dalibor Krleža |
| 6 | 03.04. | Lab2 (pandas): priprema i čišćenje podataka | |
| | 06.04. | <ul style="list-style-type: none">• Programski jezici i razvojni alati• Paketi za praktično strojno učenje• Učenje i stvaranje modela• Repozitoriji modela koji omogućuju suradnju unutar razvojnog tima• Verzioniranje modela i postupak evaluacije | Dalibor Krleža |



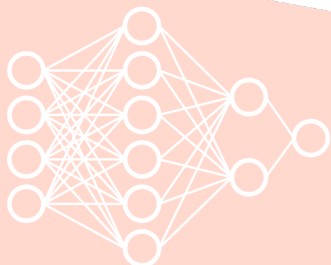
Raspored predavanja (3)

| | | | |
|---|--------|---|--|
| 7 | 13.04. | <ul style="list-style-type: none">• Mikroservisna arhitektura• Sigurnost mikroservisne arhitekture i inteligentnih servisa generalno• Integracija inteligentnih servisa u poslovne procese organizacije• Inteligentna optimizacija poslovnih procesa | Dalibor Krleža |
| 8 | 04.05. | <ul style="list-style-type: none">• Serveri za izvođenje inteligentnih servisa• Načini pokretanja modela na serveru• Verzioniranje modela pokrenutih na serveru• Skalabilnost: kontejneri i cloud• Visoka dostupnost• Sigurnost | Dalibor Krleža |
| | 05.05. | | Lab3 (MLflow): Učenje modela, repozitorij modela, suradnja unutar tima |
| 9 | 11.05. | <ul style="list-style-type: none">• Praćenje i monitoring rada inteligentnog servisa• Utjecaj rada inteligentnog servisa na podatke• Zatvorena petlja učenja, evaluacije i rada modela | |



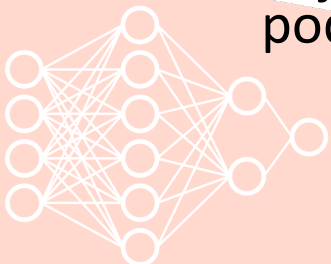
Raspored predavanja (3)

| | | | |
|----|--------|--|---|
| 10 | 18.05. | <ul style="list-style-type: none">• Metodologije za razvoj inteligentnih sustava• MLOps | Dalibor Krleža |
| 11 | 22.05. | | Lab4 (MLServer, Seldon Core 2, Prometheus): serviranje modela i praćenje rada |
| | 25.05. | <ul style="list-style-type: none">• MLOps• Izvedba pipeline-ova• Automatizacija inkrementalnog učenja modela | Dalibor Krleža |
| 12 | 01.06. | <ul style="list-style-type: none">• Cloud vs <i>on-premise</i>• Analiza efekta uvođenja inteligentnog sustava | Dalibor Krleža |
| 13 | 05.05. | | Lab5 (Airflow): pipeline, custom modeli |



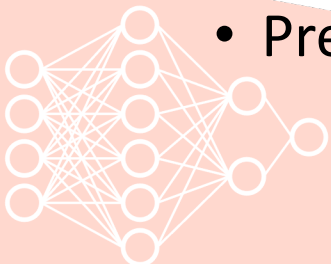
Laboratorijske vježbe (1)

- Izvođenje laboratorijskih vježbi je fizički na fakultetu
- Koristi se programski jezik Python (verzija 3)
- Laboratorijska vježba traje 180 minuta (3 sata)
 - Laboratorijske vježbe se odrađuju kroz Jupyter notebookove
 - Preporuča se imati vlastiti laptop ili barem se udružiti s nekim
 - Većina vježbi se može odraditi na standardnom laptopu od barem 4 jezgre procesora i 8 GB RAM-a
 - Jupyter notebookovi su pisani za operacijski sustav Linux
 - Preporučamo Ubuntu 22.04 LTS (desktop), može i kao virtualni stroj (recimo Oracle VirtualBox)
- Na laboratorijskoj vježbi će nas biti više kako bismo vas pružili pomoć
 - Ohrabrujemo i međusobnu komunikaciju i zajedničko rješavanje problema
 - Cilj laboratorijskih vježbi je da dobijete hand-on iskustvo s određenim temama iz područja *ML engineeringa*



Laboratorijske vježbe (2)

- Vašu radnu okolinu pripremite tako da instalirate
 - python interpreter + virtualenv
 - VS Code razvojno okruženje
 - Postoji snap za Ubuntu 22.04 LTS
 - Dodajte podršku za python i docker
 - docker
 - minikube
- Ostalo budemo u sklopu pojedine vježbe + priprema kod kuće za stvari koje duže traju
- Ponovite:
 - Predmet “Raspodijeljeni sustavi” – kontejneri i mikrousluge
 - Predmet “Strojno učenje” – algoritmi za strojno učenje



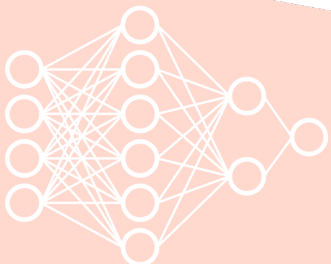
Projekt (1)

- Za potrebe vaših projekata dignuta je MLOps okolina koja se može naći na:
<https://aris.fer.hr>
- Okolina je kubernetes cluster (1.24) koja sadrži
 - PostgreSQL bazu podataka – pristupate kroz pgAdmin sučelje
 - Apache Spark
 - MLflow
 - MLServer
 - Seldon Core v2
 - Prometheus
 - Apache Airflow
 - nginx ingress
- U okolinu se prijavljujete AAI@EduHR korisničkim računom
- Zahvaljujemo True-North d.o.o. i Zoranu na pomoći s instalacijom okoline



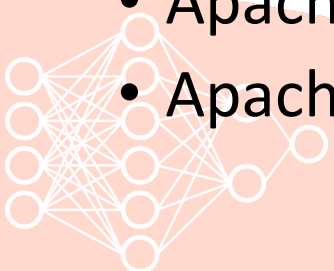
Python

- Reference Manual - <https://docs.python.org/3.10/reference/>
- Tutorial - <https://docs.python.org/3.10/tutorial/index.html>
- W3 Schools - <https://www.w3schools.com/python/>
- IDE
 - PyCharm (JetBrains) - <https://www.jetbrains.com/pycharm/>
 - Besplatna licenca za studente
 - Slično IntelliJ/CLion IDE-ovima
 - VS Code - <https://code.visualstudio.com/docs/languages/python>



Python ML

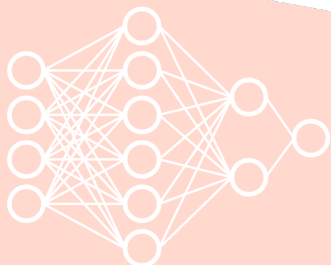
- scikit-learn: <https://scikit-learn.org/stable/>
- PyTorch: <https://pytorch.org>
- Keras: <https://keras.io>
- Tensorflow: <https://www.tensorflow.org>
- Pandas: <https://pandas.pydata.org>
- MLflow: <https://mlflow.org>
- SeldonIO MLServer: <https://www.seldon.io/introducing-mlserver>
- Seldon Core v2: <https://www.seldon.io/seldon-core-v2-the-next-generation-open-data-centric-mlops>
- Apache Spark: <https://spark.apache.org/docs/latest/api/python/>
- Apache Airflow: <https://airflow.apache.org>



Polaganje predmeta u kontinuiranoj nastavi

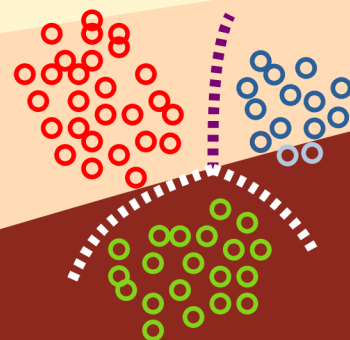
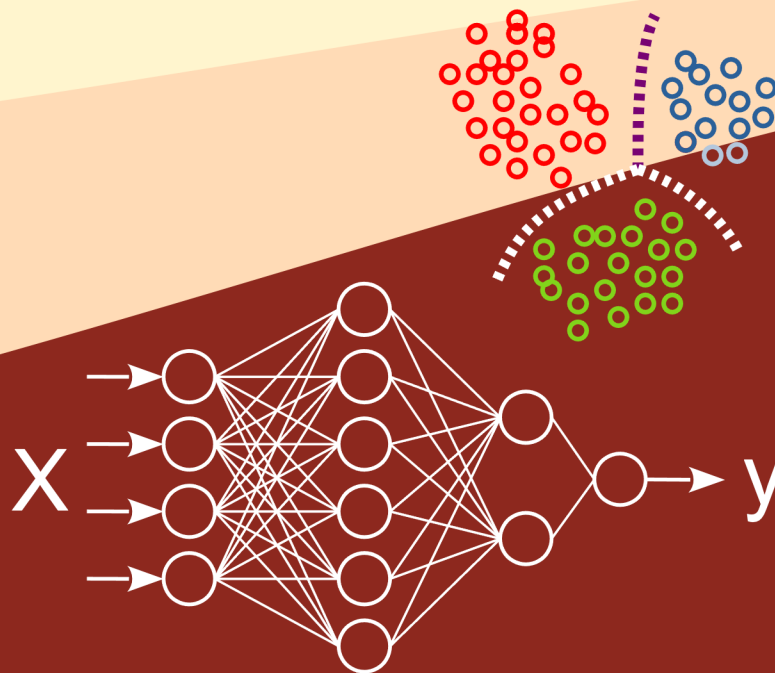
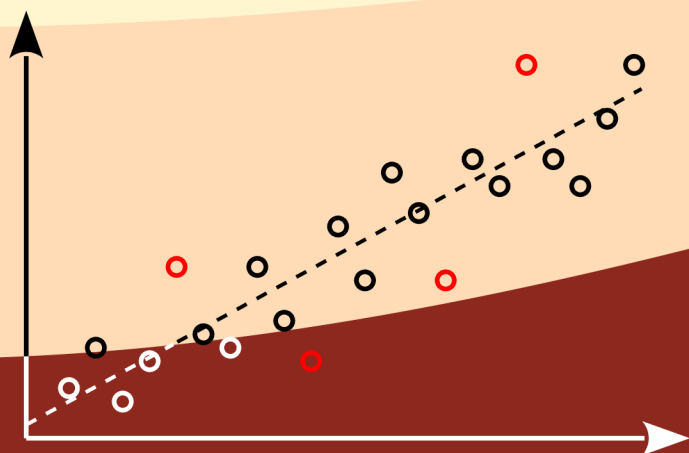
| Aktivnost | Maksimalno bodova |
|------------------------------|---------------------------|
| Laboratorijske vježbe | 20 |
| Pohađanje predavanja | 10 |
| Projekt | 70 |
| Dodatni teži zadaci (opcija) | 5 (povrh maksimalnih 100) |

- Povrh redovnih 100 bodova može se skupiti dodatnih 5 bodova rješavajući neki teži zadatak (recimo ML challenge)
- Ispitnih rokova nema. Imate projekt koji se ocjenjuje.



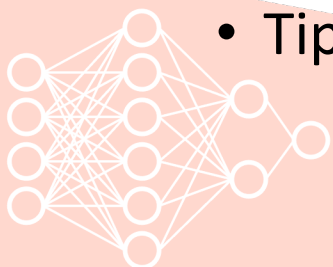
Arhitektura i Razvoj Inteligentnih Sustava

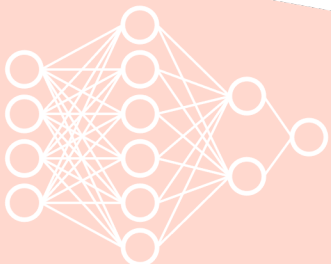
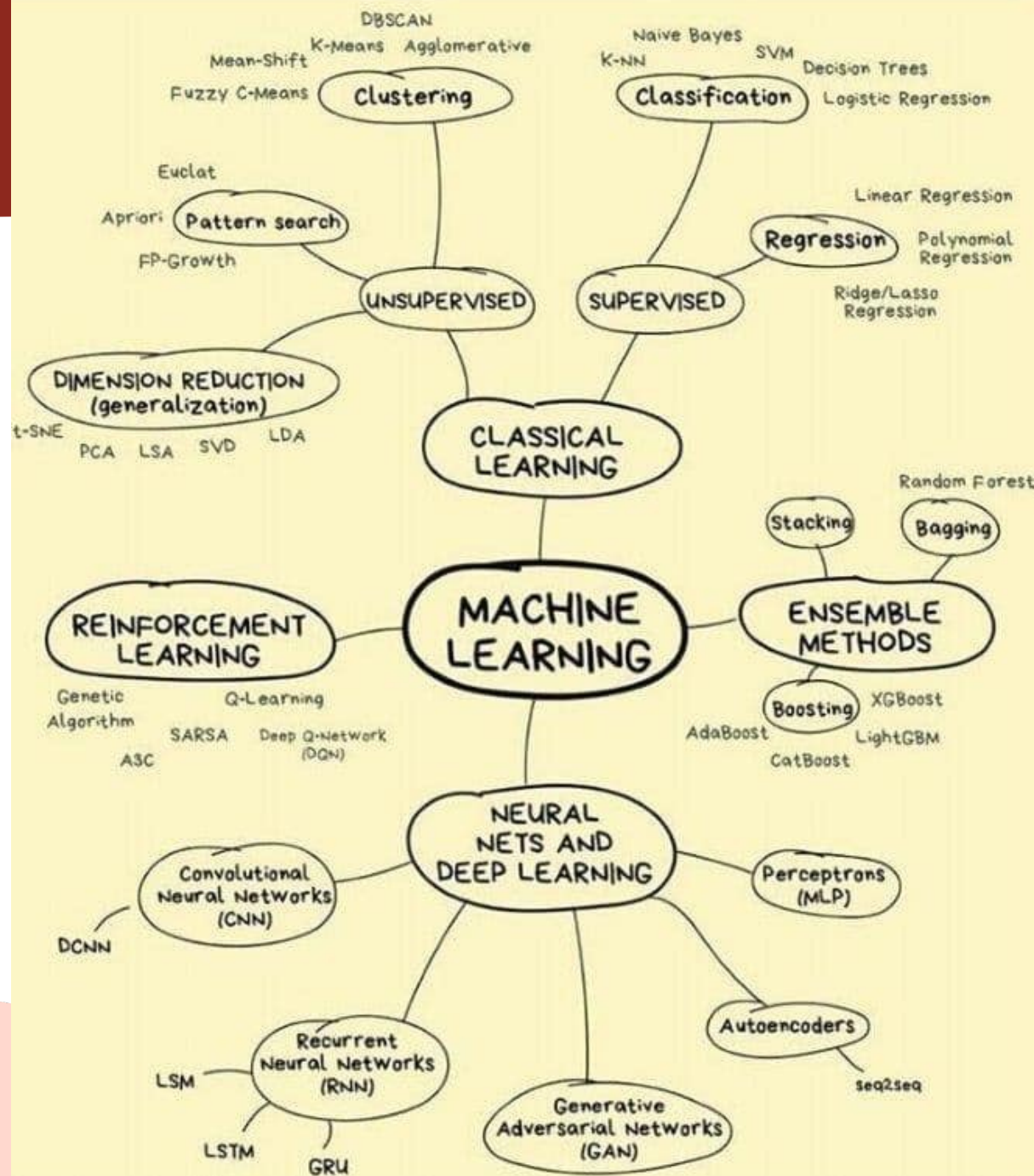
Tjedan 1: Problematika uvođenja inteligentnih sustava



Osnovi pojmovnik

- Umjetna inteligencija – Simulacija procesa ljudske inteligencije na strojevima (računala)
- Usko i generalno primjenjiva umjetna inteligencija
 - Uska se koncentrira na izvršavanje konkretnih uskih područja / zadataka
 - Generalno primjenjiva (AGI) pokušava replicirati kognitivne procese ljudskog mozga
- 4 tipa umjetne inteligencije (Arend Hintze)
 - Tip I : Repetitivni strojevi – bez memorije, ponavljanje naučenog
 - Tip II : Limitirana memorija – strojevi koji imaju limitiranu memoriju kod donošenja odluke
 - Tip III : Theory of mind – Uključeno emocionalno stanje
 - Tip IV : Samosvijest





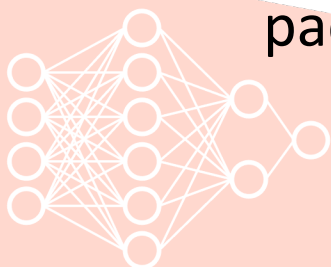
Primjene umjetne inteligencije

- Automatizacija (*automation*) – skup izvršnih elementa i senzora sa pozadinskim pravilima koja omogućuju izvršenje nekog zadatka – proizvodna linija
- Strojno učenje (*machine learning*) – spremanje znanja u parametre nekog algoritma, a koje algoritmu omogućuje reprodukciju tog znanja
 - Nadzirano učenje – labele su zadane unaprijed
 - Nenadzirano učenje – labele nikad nisu zadane
 - Podržano učenje – labele nisu zadane, ali se model nagrađuje ili kažnjava u trenutku dobre ili loše odluke
- Strojni vid (*machine vision*)
- Procesiranje prirodnog jezika (*natural language processing*)
- Robotika (*robotics*)
- Autonomna vozila (*self-driving cars*)
- Agregacije – Chat GPT
- U sklopu predmeta se nećemo baviti autonomnim vozilima, robotima i AGI-em generalno, već usko primjenjivim inteligentnim sustavima za konkretne zadatke i slučajeve korištenja



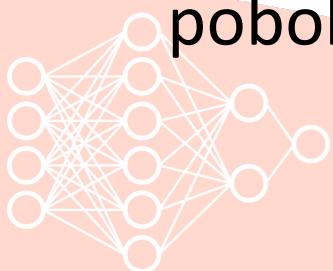
Inteligentni sustav

- Pojam „ekspertni sustav” – ljudsko znanje preneseno u računalni program, pomažući tako u poslovanju organizacije
- Pojam „znanja” postaje široko definiran
 - Osim ljudskog znanja, tu je i znanje koje se može izvući iz podataka
 - Organizacije tijekom poslovanja skupljaju velike količine podataka
 - Senzori
 - Poslovni sustavi
 - Vanjski skupovi podataka (tipični obrasci u industriji)
- Sinergija podataka i ljudskog znanja
 - Primjer je zdravstvo – velika količina biomarkera i podataka sa senzora pacijenta koji se kombiniraju sa znanjem liječnika specijalista

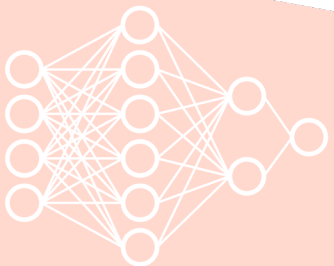
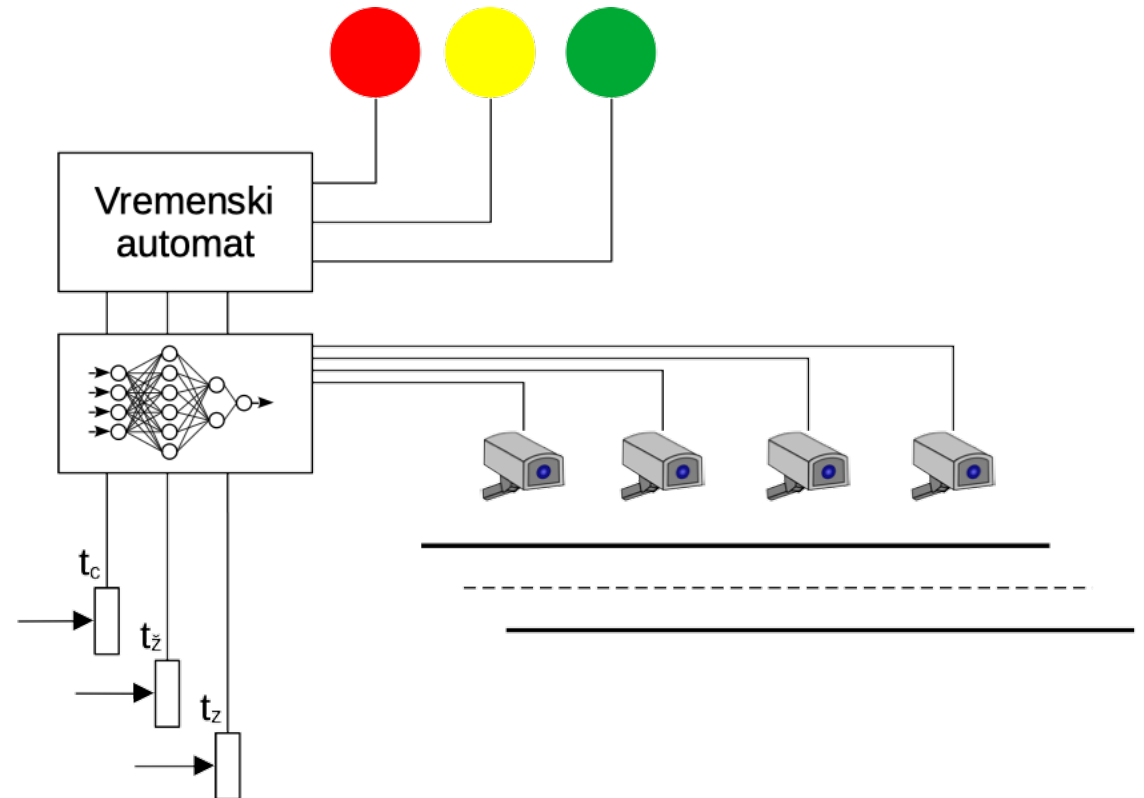
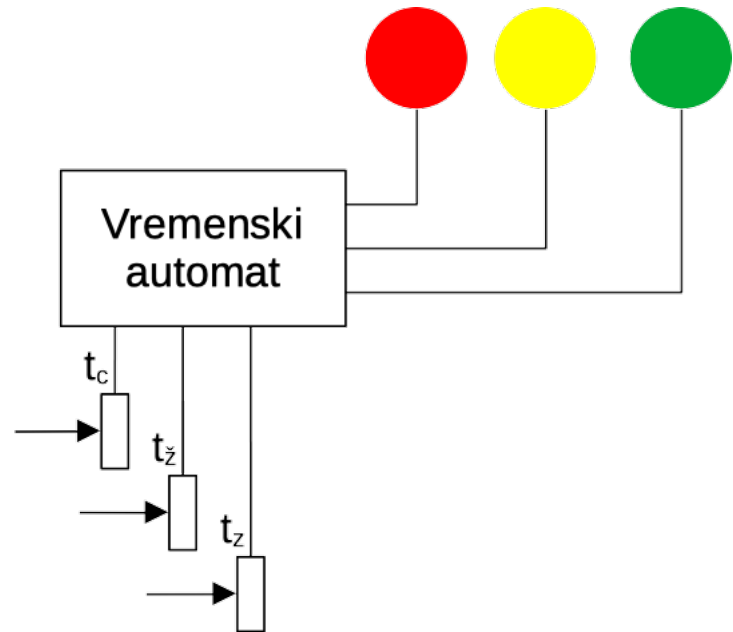


Inteligentni sustav

- U svijetu senzora, kamera, pametnih uređaja i strojeva treba postojati centralna komponenta koja povezuje ulaze i izlaze
 - Automatizirano – skripte posebno pisane u tu svrhu
 - Sustav koji se dobio učenjem?
- U svijetu poslovnih informacijskih sustava, nadogradnja koja omogućava informacijskom sustavu koji reagira na inteligentan način
 - Uočavanje obrazaca i pozadinska logika koja ima za zadaću:
 - Povećati zadovoljstvo korisnika i klijenata
 - Učiniti (poslovne) procese sigurnijim i efikasnijim
 - Smanjiti troškove
- Centralna komponenta koja sadrži logiku (inteligentnu) za značajno poboljšavanje rada okolnih sustava

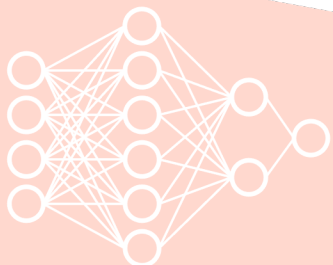
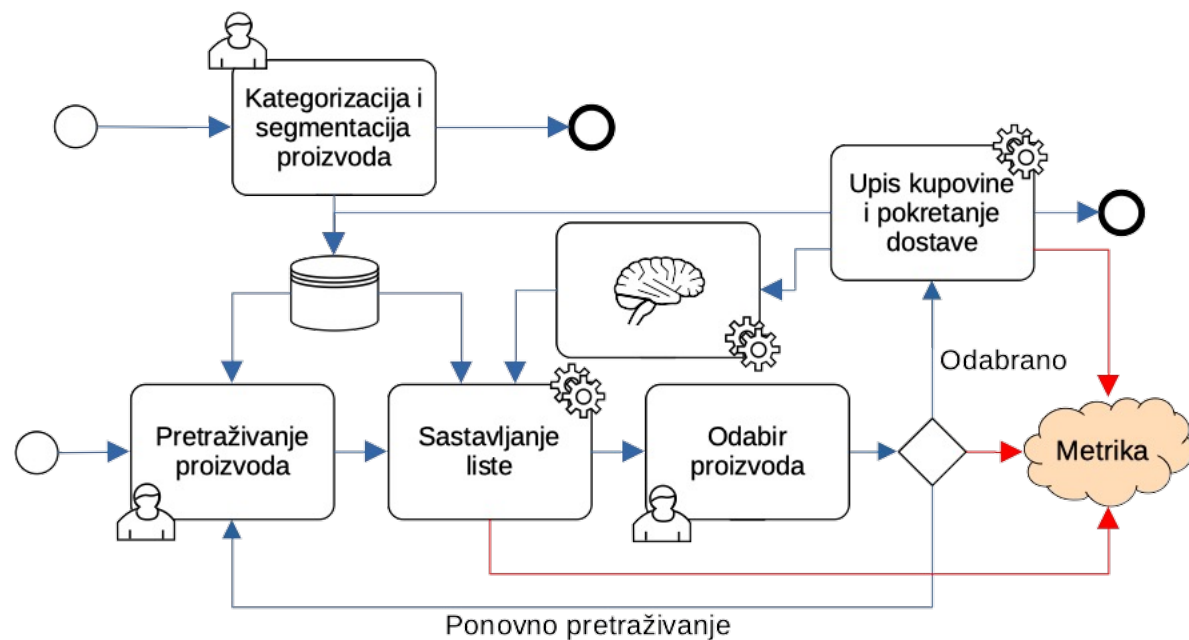
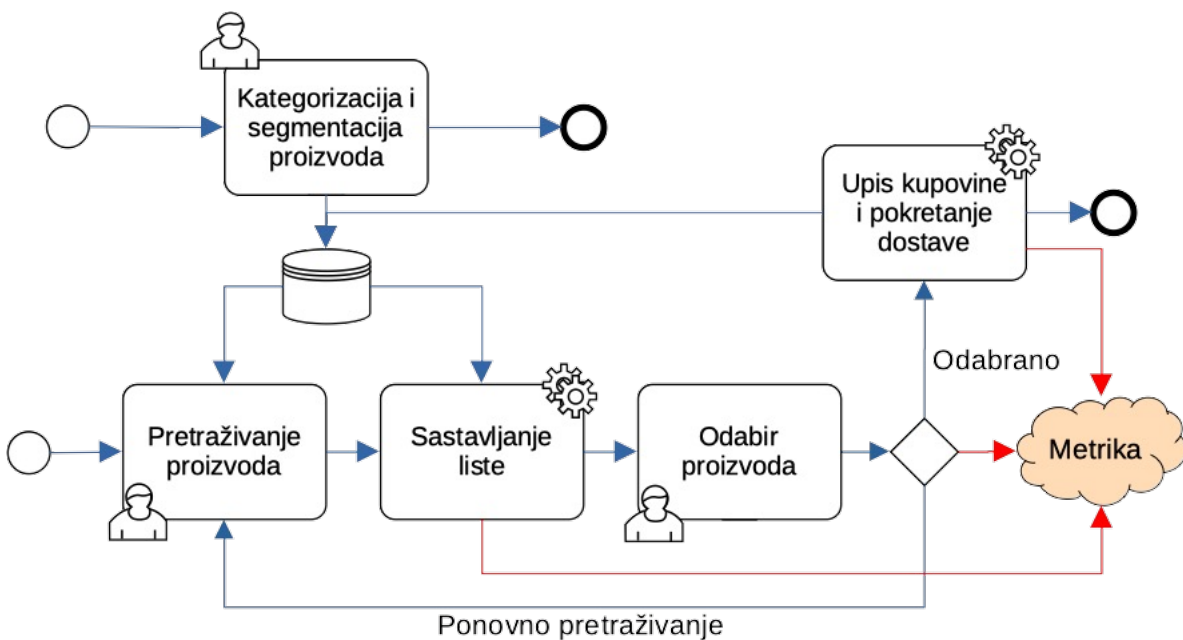


Intelligentni sustav



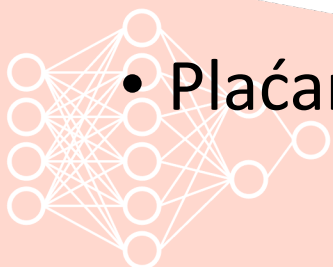
[Embedded CNN](#)

Inteligentni sustav



Gdje su svi ti inteligentni sustavi?

- U današnje vrijeme AI hype-a nekako ga teško uočavamo u svakodnevnom životu... ?
 - Izuzmimo očite primjere: Chat GPT, Photomath, biometrijske prijave, preporučivač (*recommender*) na Amazonu npr..., neke eksperimentalne primjene poput generiranja fotografija, slika, glazbe, ...
- Gdje je sav taj AI u našem svakodnevnom životu? Zašto nema očitih i značajnih primjera u našoj okolini?
 - Zar je stvarno tako teško uvesti AI u informacijske sustave?
 - Da li bi AI u sustavima koji nas okružuju trebao biti uočljiv? „Da, tu se zbilja vidi da postoji nekakva inteligentna logika u tom sustavu!”
 - Kad pitamo to pitanje, često se čuje „Pokušali smo, ali....”



- Plaćanje parkinga ispod Kvaternikovog trga

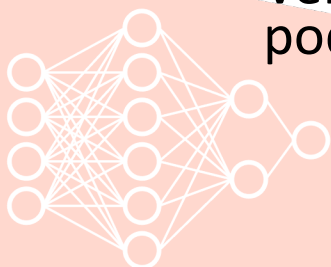
Problem očekivanja

- "Pa to je samo malo nekih tamo if...then...else naredbi"
- Sales velikih IT kompanija često u svojim prezentacijama i radionicama maskira kompleksnost područja – puno klikanja nakon kojeg nešto tamo u cloud-u prepozna da je ovo stol, a ovo pas, a ovo mačka, a ovo čaša na stolu
- Poduzeća teško odgovaraju na pitanje slučaja korištenja
 - Slučaj korištenja je banalan i nije za AI – uvodi se samo radi hype-a
 - Ili slučaj korištenja nije na mjestu koje će generirati najveću korist
- Često ispada da je korisnost ispod troškova uvođenja
 - Rezanje troškova uvođenja često rezultira neuspjehom uvođenja
 - „Ma zar mi tu zbilja moramo dodati toliko infrastrukture”
- Nema ozbiljne namjere uvođenja
 - „Samo malo isprobavamo da li je to za nas”
 - Prekratak rok
 - „Ma to su neki tamo sustavi o kojima čitamo na Internetu”
- Promjena koja rezultira kratkoročnom negativnom metrikom
- Područje u kojem imamo brze socio-ekonomske promjene, a koje se ne stignu pratiti
 - Konstantni model drift koji se ne stigne pratiti adekvatno
- Nikad niste zapravo završili, uvijek se treba nešto prilagođavati, učiti, isprobavati, provjeravati



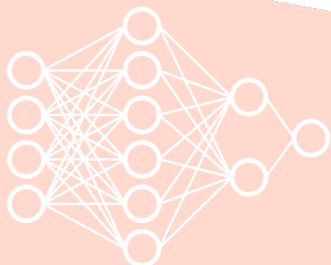
Problem podataka

- Nema dovoljno podataka za uvođenje – jednostavno ne postoji ulaz na temelju kojeg se može skupiti dovoljno znanja
- Podaci su neadekvatni
 - Puno kategoričkih podataka, premalo kontinuiranih – diskretizacija koja ne doprinosi
 - Podaci nisu ažurni i ne skupljaju se redovito
 - Puno podataka fali
 - Separacija učenje / validacija / test
- Podatke nije moguće skupiti na jednostavan način
 - Često problem ugrađenih sustava – iz kojih je teško izvaditi i spremiti podatke senzora
 - Razna nekompatibilna sučelja i nemogućnost transformacije podataka
- Nepostojanje stručne interpretacije
 - Velika količina podataka nema labela i ne postoji stručna osoba koja bi ocijenila te podatke



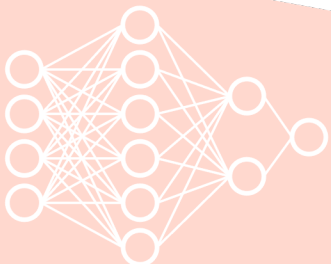
Problem podataka

- Jaki drift modela koji se ne može ispraviti s dostupnim podacima
 - Primjer IoT zgrade
- Visoka dimenzionalnost podataka
 - Razbijanje na segmente, puno malih modela koji se teško koriste
- Nepostojanje algoritma koji bi se mogao primijeniti na podatke
 - Nemogućnost prilagodbe podataka na algoritme



Problem infrastrukture

- Zastarjela infrastruktura
 - Stari izvori podataka (hijerarhijske baze podataka npr...)
 - Nepostojanje prave servisno orijentirane arhitekture – stari *green* i *gray* ekrani
- Iako je presjek infrastrukture između klasičnih informacijskih sustava i inteligentnih sustava značajan
 - Trošak uvođenja može značiti značajnu investiciju u infrastrukturu
 - Promjene u infrastrukturi mogu zahtijevati i zahvate u postojeće klasične informacijske sustave (sigurnost, servisno, ...)



Problem developera

- Svi mi danas znamo MLOps?
 - Ili možda mislimo da znamo?
 - Većina developera zna nacrtati faze MLOps-a i napisati Open Source produkte koji ga podržavaju
 - Dosta developera je sudjelovala u nekim izazovima
 - Uzmemo skup podataka s Kaggle-a, pročistimo, napravimo neku ML arhitekturu, nahranimo ju s podacima i vidimo izlaznu metriku
 - Pitanje "ML scientist" vs „ML engineer"
 - Neki developeri su išli i na pokoji tečaj
 - Takvi tečajevi, posebno u sklopu većih korporacija, završe kao sales-pitch za njihovu okolinu (IBM, RedHat, ...)
 - Iz takvih tečaja se dosta teško nauče detalji „ispod haube"

