

Izrada projekta

Sistemi za istraživanje i analizu podataka

Rokovi

Prijava timova i (okvirnog) naslova teme	10.11.2024.	Na sajtu predmeta biće objavljen <i>google sheet</i> u koji treba da se upišete
Predaja predloga projekta	24.11.2024.	Asistentu na mail
Rok do koga ćete dobiti komentare za ispravku	08.12.2024.	\
Predaja revidiranog predloga projekta	22.12.2024.	Asistentu na mail
Rok do koga dobijate notifikaciju da li je projekat prihvaćen	29.12.2024.	\
Prva kontrolna tačka	Okvirno 20.01.2025.	Usmene konsultacije sa asistentom u zakazanom terminu
Druga kontrolna tačka	Okvirno 20.02.2025.	Usmene konsultacije sa asistentima i profesorom u zakazanom terminu
Predaja finalne verzije projekta	20.03.2025.	Asistentu na mail

Asistenti

- Katarina-Glorija Grujić (katarina.glorija@uns.ac.rs)
- Aleksandar Vujinović (aleksandar@uns.ac.rs)
- Mihaela Osmajić (mihaela.osmajic@uns.ac.rs)
- Nakon prijave teme i timova, javićemo vam koji asistent vodi vaš projekat

Izrada projekta

- **Timski rad** u grupama od 2-3 studenata
 - Svaki student mora da da svoj doprinos – tražićemo da specificirate ko je šta radio
 - Usled velikog broja studenata, morate raditi u timu
 - Ako ne možete da pronađete tim
 - Javite se asistentima *pre* roka za prijavu timova da vam pomognu da pronađete tim
 - Naznačite interesovanja/oblasti/teme koje su vam prihvatljive
- Tim samostalno istražuje i predlaže temu
 - Dobijate sugestije gde možete da pronađete skupove podataka i da se informišete o temama koje vas zanimaju: [*Izrada projekta/Predlog projekta – uputstvo.pdf*](#)

Predaja predloga projekta

- Kratak (ali informativan) izveštaj na 1-3 strane A4 formata
- Šaljete asistentu koji vodi vaš projekat na mail do propisanog roka
- Sugestije kako sastaviti predlog projekta se nalaze na sajtu predmeta u [/Izrada projekta/Predlog projekta – uputstvo.pdf](#)
- Primere predloga projekata možete naći na sajtu predmeta u [/Predlozi projekta prethodnih generacija](#)
- Na sajtu predmeta ćete naći predloge projekata od starijih generacija. Molimo Vas da ovo pažljivo pregledate – [nećemo prihvatati predloge projekta koji su identični onima iz stare generacije!](#)

Predlog projekta – OBAVEZAN sadržaj

1. Jasna definicija cilja projekta
2. Motivacija problema rešavanog u projektu
3. Relevantna literatura (minimum 3 rada)
 - Svaki rad mora biti opisan sa svih 6 obaveznih elemenata: (1) cilj rada, (2) opis metodologije, (3) opis skupa podataka, (4) način evaluacije (postupak i mera), (5) rezultati i (6) **zaključak – zašto je ovaj rad relevantan za vaš projekat**
4. Skup podataka
 - Postojeći skup:
 - (1) link ka skupu podataka, (2) ciljno obeležje i njegov sadržaj (3) atributi
 - Sami sastavljate skup:
 - (1) koje attribute dobavljate, (2) odakle, (3) šta je ciljno obeležje i šta ono sadrži, (4) kako ćete dobiti labela
5. Predložena metodologija
 - **Mora** biti bazirana na radovima navedenim u relevantnoj literaturi
6. Metod evaluacije: (1) postupak i (2) mera evaluacije
7. (Opciono) softver koji planirate da koristite
 - Detaljniji opis je u dokumentu *Predlog projekta – uputstvo.pdf*

Predlog projekta – NE smete

- Prekršiti zadate rokove
 - U suprotnom, projekat je automatski odbijen i morate raditi predefinisani projekat za maksimalnu ocenu 6.
- Tražiti od asistenata da vam pronađu temu ili razrađuju detalje projekta
 - Ovakve stvari nas eventualno možete usmeno pitati, ali na mailove sledeće sadržine ne odgovaramo:
 - Da li je ovo ok skup podataka <link/opis skupa podataka>
 - Da li je ovo ok tema <naslov teme u par reči>
 - Da li mi možete predložiti skup podataka za...
 - Odgovor je ne znamo – odgovor na sva ova pitanja zavisi od toga koji je vaš cilj, metodologija... i ostalo što vas teramo da specificirate u predlogu 😊
 - Da vam odgovorimo, i mi bismo morali pretraživati literaturu kao i vi, a to je ipak deo vašeg zadatka
 - Zato ćemo na pitanja da li je tema/skup podataka/cilj projekta... u redu ili nije, odgovoriti *isključivo* kada nam pošaljete *kompletan* predlog (sa svim obaveznim sadržajem)

Predlog projekta – NE smete

- Slati nepotpun predlog projekta (videti obavezan sadržaj)
 - Ako je rok istekao – projekat je automatski odbijen
 - Ako rok nije istekao – sve što ćemo uraditi je da vas podsetimo da do roka pošaljete kompletan predlog
- Tražiti pomoć a da ne znate osnovne pojmove sa predavanja, npr.
 - Šta je nadgledano (*supervised*) a šta nenadgledano (*unsupervised*) učenje
 - Šta je klasifikacija/regresija/klasterovanje
 - Kada je primenljivo klasifikacija/regresija/klasterovanje (u zavisnosti od postojanja i tipa ciljne labele)
 - Koji modeli se mogu primeniti u slučaju klasifikacije/regresije/klasterovanja
 - Sve ovo ćete čuti na nastavi, molimo vas da dolazite

Predlog projekta – Smete (i poželjno je)

- Da pre isteka roka pošaljete asistentu predlog
 - Ograđujemo se da primamo *Isključivo kompletne predloge*
- Javićemo vam:
 - Da li je ideja prihvatljiva ili ne
 - Kako da ispravite ili dopunite projekat (ako je ideja prihvatljiva i ako ste naveli sve neophodne elemente)
- Na ovaj način, obezbeđujete se (pre isteka roka) da će vaš projekat biti prihvaćen

Dileme oko sastavljanja predloga projekta

- Najlakši način da ih rešite jeste da potražite radove koji su koristili skup podataka koji ste odabrali (ili neki veoma sličan) i da vidite šta je u njima rađeno:
 - Šta je problem koji rešavaju i koja je motivacija za to
 - Kako su došli do ciljne labele
 - Na osnovu čega su predviđali/klasterovali/... (koje attribute su koristili)
 - Koje tehnike su primenili

Prva kontrolna tačka – šta treba da uradite

- U ovom trenutku očekujemo da ste krenuli sa izradom projekta
 - Imate skup podataka na kome radite
 - ako ga sami kreirate, treba da ste ga formirali barem delimično tako da možete da isprobate neki osnovni model koji ste naveli u predlogu
 - Odradili ste detalju analizu podataka
 - Imate neke osnovne rezultate da nam pokažete
- To može biti primenjen neki osnovni algoritam za klasifikaciju/regresiju/klasterovanje...
 - Rezultati ne moraju biti dobri
 - Parametri ne moraju biti optimizovani

Imate neku meru performansi tog algoritma (npr. *accuracy*, R^2 , analizirali ste klustere, ...)

Prva kontrolna tačka – cilj

- Da pokažete da je projekat koji ste zamislili izvodiv
- Da pokažete da ste se upoznali sa problemom koji rešavate i imate jasno definisan cilj
 - Npr. ako se radi o predikciji neke vrednosti – šta je ciljna varijabla, koje vrednosti može da uzima, kako su te vrednosti zastupljene u skupu podataka, itd.
- Da steknete osećaj koliko je problem težak
 - Imate neki *baseline* (osnovne rezultate)
 - Odatle vidite koliko ste na početku izrade daleko od očekivanih rezultata (iz literature)
 - Ako je tačnost već sada jako visoka (problem je prelak), treba da smislite kako treba promeniti ili dopuniti projekat da imate mogućnosti za unapređenje i visoku ocenu

Prva kontrolna tačka – kako branite

- Radi se o usmenoj odbrani kod asistenta u zakazanom terminu
 - Ponesite fajlove koji su vam potrebni da pokažete sve što ste odradili (možete i na sopstvenom laptopu)
 - Odbrana je obavezna – ako iz opravdanih razloga niste u mogućnosti da dođete u zakazanom terminu, molimo vas da **pre isteka roka** javite da zakažemo drugi termin
 - Dovoljno je da prisustvuje jedan član tima
- Ako ne odbranite prvu kontrolnu tačku do propisanog roka, konačna ocena projekta vam se na kraju umanjiti za jednu ocenu

Druga kontrolna tačka – šta treba da uradite

- U ovom trenutku očekujemo da ste prilično napredovali sa projektom
 - Izvršili eksplorativnu analizu podataka
 - Isprobali modele koje ste naveli u predlogu
 - Optimizovali ih
 - Možete da izvedete neke zaljučke iz svojih rezultata
- Gotovo sve što ste obećali u predlogu projekta treba da bude urađeno

Druga kontrolna tačka – cilj

- Da konkretizujemo
 - Koju ocenu možete dobiti sa trenutnim stanjem projekta
 - (Ako je potrebno) kako da doradite projekat da biste dobili najvišu ocenu

Druga kontrolna tačka – kako branite

- Radi se o usmenoj odbrani kod asistenta u zakazanom terminu
 - Ne morate spremiti prezentaciju, ali
 - ponesite fajlove koji su vam potrebni (možete i na sopstvenom laptopu)
 - ne zaboravite da izložite asistentima sve šta ste uradili – imajte podsetnik, beleške,...
 - Odbrana je obavezna – ako iz opravdanih razloga niste u mogućnosti da dođete u zakazanom terminu, molimo vas da **pre isteka roka** javite da zakažemo drugi termin
 - Dovoljno je da prisustvuje jedan član tima
- Ako ne drugu kontrolnu tačku ne odbranite do propisanog roka, konačna ocena projekta vam se na kraju umanjiti za jednu ocenu
 - Druga kontrolna tačka se smatra odbranjenom ukoliko asistenti procene da ste uradili dovoljnu količinu posla iz predloga projekta

Predaja završnog izveštaja

- Potrebno je da do naznačenog roka asistentu na mail pošaljete završni izveštaj
 - Radi se o tekstualnom izveštaju u kome opisujete
 - Problem koji ste rešavali i motivaciju za njegovo rešavanje
 - Pregled relevantne literature koju ste pročitali
 - Skup podataka
 - Metod/algoritme koje ste primenili
 - Rezultate
 - Zaključke
- Na sajtu predmeta su okačene i detaljnije sugestije o sadržaju izveštaja [/Izrada projekta/Pisanje završnog izveštaja o projektu.pdf](#)
- Na sajtu predmeta se nalazi i propisan format (IEEE Template u dve kolone) [/ Izrada projekta/Sablon_zavrsni_izvestaj](#)
- Izveštaj mora biti na srpskom jeziku
- Propisan obim: 6-8 strana za grupe od 2 člana i 8-10 strana za grupe od 3 člana

Predaja finalne verzije projekta

- Ako izveštaj završite pre isteka roka, možete ga poslati asistentu na mail kako biste dobili komentare o samom tekstu izveštaja
 - Pravo na ovo imate samo ako ste ispoštovali i prvu i drugu kontrolnu tačku
 - Ovo možete uraditi u najviše dva navrata. Zato se potrudite da kada dobijete sugestije ispravite što je moguće više, kako biste dobili još konstruktivnih komentara u narednoj iteraciji
- Primere lepo urađenih projekta možete naći na sajtu predmeta u [/Izrada projekta/](#)
 - [Primer dobrog projekta 1.pdf](#)
 - [Primer dobrog projekta 2.pdf](#)

Šta se sve ocenjuje

- Poštovanje obaveza u zakazanim rokovima
- Koliko ste dobro upoznati sa problemom
 - Da li je literatura koju ste izložili dovoljno obimna i relevantna
 - Korisne sugestije za čitanje naučnih radova se nalaze na slajdovima */Izrada projekta/ Kako procitati naucni rad.pdf*
 - Nije dovoljno da u konačnom izveštaju pobrojite samo onu literaturu koju ste naveli u predlogu projekta. Očekujemo da ste u međuvremenu pročitali mnogo više radova
- Koliko toga je urađeno (timski i pojedinačno)
 - Procenu šta sve treba da uradite/doradite za maksimalnu ocenu dobićete na kontrolnim tačkama
 - Pored toga, u svakom trenutku (do predaje finalnog izveštaja) možete dogovoriti konsultacije kako biste proverili da li projekat ide u pravom smeru
- Koliko dobro je napisan završni izveštaj
 - Mora da bude stilski dobro napisan
 - Mora da poštuje propisan šablon
 - Mora da bude adekvatnog obima
 - Svaki deo izveštaja mora da bude izložen na adekvatan i razumljiv način (detaljan podsetnik možete naći na sajtu predmeta *Pisanje završnog izveštaja o projektu.pdf*)
 - Literatura mora biti dovoljno obima i citirane rečenice moraju biti razumne i dobro povezane.

Česti propusti u izradi projekta

- Niste obrazložili sve izbore parametara. Na primer,
 - zašto takva podela na trening/test skup
 - zašto ta mera evaluacije
 - zašto smatrate da je rejting od 3 pa naviše pozitivan, itd.).
- Moguća obrazloženja:
 - do vrednosti parametara ste došli empirijski (optimizacijom modela)
 - prilikom eksplorativne analize ste uočili neki šablon
 - na osnovu domenskog iskustva
 - drugi autori koriste slične postavke parametara/način i meru evaluacije...
- Primenili ste model a zaboravili da optimizujete vrednosti hiperparametara
- Trenirali ste model ili optimizovali parametre na istim podacima (ili na podskupu podataka) na kome evaluirate model. Napravite razliku između trening/validacionog/test skupa

Česti propusti u izradi projekta

- Ne koristite adekvatnu meru performansi
 - Na primer, tačnost nije dobra mera u slučaju da klase skupa podataka nisu balansirane
- Ako koristite veliki skup podataka (imate problem sa memorijskim resursima):
 - *RapidMiner* verovatno nije dobro rešenje!
 - Razmislite o primeni *big data* tehnika i drugih softvera/programskih jezika
 - Razmislite da li ima smisla grupisati podatke u smislene podskupove i kreirati odvojene modele

Moguća poboljšanja kvaliteta projekta

- „Dodatne poene“ možete osvojiti analizom grešaka modela. Izdvojite podskup primera na kojima je vaš model pogrešio i pokušajte da razumete razlog
 - Npr. napravili ste model za automatsko prepoznavanje sentimenta teksta. Po inspekciji grešaka koje model pravi utvrdili ste da ne prepoznaje negaciju ili sarkastične komentare
 - Za detaljnije instrukcije o načinu analize grešaka modela preporučujem <https://github.com/ajaymache/machine-learning-yearning>, poglavlja 14 i 15
- Razmislite da li možete da kombinujete postojeće pristupe, npr. da li možete da kombinujete glasove različitih klasifikatora
- Razmislite da li možete kombinovati skup podataka sa dodatnim izvorima podataka
- Ako nemate dovoljno podataka, razmislite da li možete da koristite *semi-supervised learning* ili *active learning*

Dodatne napomene

- Možete biti pozvani da usmeno tumačite predat izveštaj, a pogotovo ako posumnjamo da je izveštaj generisan od strane AI alata (npr. ChatGPT)
- Ocena može biti automatski smanjena ukoliko izveštaj ne ispunjava sve prethodno navedeno.
- Teme vezane za sport i predikciju rezultata sportskih utakmica nisu dozvoljene

Primeri predloga projekata

(prethodne generacije studenata)

Video igre

- Predikcija ishoda meča u igrici Dota2 i predlog heroja sa kojim igrač ima veće šanse za pobedu
 - RPG igra u kojoj su igrači podeljeni na dve frakcije
 - Na početku meča igrači biraju jednog (od preko 100 mogućih) heroja, pri čemu heroji imaju neke svojstvene karakteristike (strength, agility, intelligence)
 - Motivacija: nagradni fond za takmičenje *The International Dota 2 Championships* je preko 18 miliona dolara



Capital bikesharing

- Cilj: predviđanje lokacija novih stanica za bicikle i uklanjanje nepotrebnih stanica
- Motivacija:
 - Ekspanzija sistema za iznajmljivanje bicikala u velikim gradovima
 - Stanice sa većim prometom su popularnije i donose veći prihod – uvećenje profita i zadovoljstva korisnika
 - Za otvaranje ili zatvaranje stanice kompanija mora da uloži finansijska sredstva



Capital bikesharing

- Skup podataka
 - <http://www.capitalbikeshare.com/trip-history-data> - informacije o trajanju putovanja (datum, vreme početka i kraja putovanja), početna i krajnja stanica (i geografske koordinate), tip članarine,... Takođe se na osnovu datih podataka može odrediti popularnost stanica (ciljna varijabla)
 - <http://wiki.openstreetmap.org/wiki/> - preuzeti podaci o objektima u blizini stanica: stanice metroa, velike kompanije, znamenitosti, restorani,... Uticaj objekata je rangiran prema udaljenosti od stanice (u minutima šetnje i km)
 - Podaci su ručno filtrirani i pročišćeni
- Problem je definisan kao problem klasifikacije – stanice su razvrstane u 5 kategorija popularnosti
- Primenjen je model stabla odlučivanja
- Evaluacija: unakrsna validacija, merenje preciznosti i odziva

Sentiment analiza tweetova

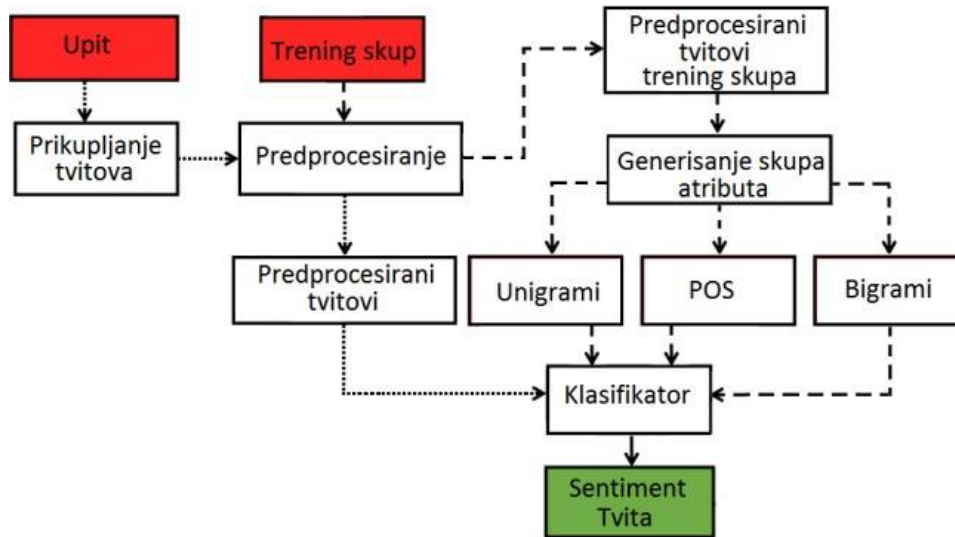
- Cilj: automatsko određivanje sentimenta tweet-a (pozitivno ili negativno)
- Motivacija:
 - Automatsko određivanje mišljenja ciljne grupe korisnika o događajima, proizvodima, poznatim ličnostima, kompanijama, ...
 - Primena u društvenim naukama (sociološke, ekonomske, istorijske, pravne), novinarstvu i reklamnim kampanjama
 - Kompanije mogu na ovaj način da ispituju javno mnjenje o svom proizvodu
- Skup podataka:
 - Sakupljani su tweet-ovi kojima je automatski dodeljivan sentiment na osnovu emotikona
 - Za potrebe evaluacije modela korišćen je ručno anotiran skup podataka

<i>Emotikon</i>	<i>Osećaj</i>
:D, 8-D, 8D, x-D,xD, X-D, XD, =-D, =D	very happy
:), :, :D, :o), :], :3, :c), :>, =], 8), =), :}, :^), :>)	happy
;-), :), *-), *), :~], :], :D, :^), :-,	wink
>:P, :-P, :P, X-P, x-p, xp, XP, :-p, :p, =p, :-P, :P	cheeky
>:[, :-(, :(, :-c, :c, :-<, :>C, :<, :-[, :[, :{	sad



Sentiment analiza tweetova

- Formiran je pipeline za procesiranje tweet-ova



- Ispitivane su performanse različitih klasifikacionih algoritama (NB, SVM, Random forest,...) u kombinaciji sa različitim obeležjima skupa podataka (bez/sa POS tagova, formiranih bigrama,...)

Sentiment analiza – leksički pristup

- Leksički pristup podrazumeva postojanje unapred pripremljenog rečnika u kojima je dostupan polaritet reči (reči su anotirane kao pozitivne, negativne i neutralne na osnovu toga koliko se često pojavljuju u odgovarajućim kontekstima)
- Postoje mnogi ovakvi rečnici:
 - Christopher Potts Sentiment Tutorial 2011 <http://sentiment.christopherpotts.net/lexicons.html>
 - The General Inquirer <http://www.wjh.harvard.edu/~inquirer/>
 - LIWC (Linguistic Inquiry and Word Count) <http://liwc.wpengine.com/>
 - MPQA Subjectivity Cues Lexicon http://mpqa.cs.pitt.edu/lexicons/subj_lexicon/
 - Bing Liu Opinion Lexicon <https://www.cs.uic.edu/~liub/fbs/opinion-lexicon-english.rar> (njihova stranica na temu Opinion Mininga <https://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html>)
 - SentiWordNet <http://sentiwordnet.isti.cnr.it/>
- Obično se u leksičkom pristupu polaritet tvita određuje nekom kombinacijom polariteta pojedinačnih reči
- Različiti rečnici mogu da daju veoma različite rezultate
- [Moguće je i kombinovati više različiti rečnika. Na primer, u radu http://people.mpi-inf.mpg.de/~smukherjee/research/cikm2012-twisent.pdf je polaritet tvitova određivan većinskim glasanjem 4 različita leksikona](http://people.mpi-inf.mpg.de/~smukherjee/research/cikm2012-twisent.pdf)
- Možete ovaj pristup kombinovati sa nadgledanim obučavanjem

Sentiment analiza

- Predikcija rejtinga restorana/usluge/servisa na osnovu sentimenta komentara
 - Xu, Y., Wu, X. and Wang, Q., 2015. Sentiment Analysis of Yelp's Ratings Based on Text Reviews – korišćeni su komentari sa Yelp-a
 - Dragan Vidaković
 - Sakupljeni komentari sa sajta <http://www.donesi.com/>
 - Korisnički nickname, naslov komentara, vreme postavljanja, tekst komentara, opis restorana, rejting,...
 - Ručno čišćenje komentara (ćirilica, ch-> č, non-ASCII),...
 - Izazov: nema puno resursa za srpski jezik
 - Eksplorativna analiza
 - Po zemljama
 - Kako izgleda prosečan pozitivan/negativan komentar (pridevi, često korišćene reči, velika/mala slova, dužina rečenica, interpunkcija,...)
 - Koji atributi će imati uticaja,...
 - Slična ideja mogla bi se sprovesti na nekom drugom sajtu, npr. <http://oceniprofesora.com/>

Sentiment analiza

- Slično, možete raditi analizu sentimenta revizija filmova, novinskih članaka (i komentara novinskih članaka), blogova, youtube komentara,...
- Npr. predviđanje rejtinga osnovu komentara



- Automatsko sortiranje muzike po žanrovima
 - Motivacija: razvoj softvera za asistenciju prilikom izbora muzike

Text mining

- Prepoznavanje plaćenih komentara
- Skinuti su komentari sa B92 i Blica i anotirani od strane korisnika kao “bot” ili “nije bot” <http://startit.rs/lovac-na-sendvice-bot/>
- Izazov: ne postoji mnogo alata za procesiranje srpskog jezika



Sentiment analiza

- IEMOCAP baza podataka <http://sail.usc.edu/iemocap/>
 - Radi se o detekciji emocija (bes, sreća, gađenje,...) na osnovu audio-vizuelnih podataka
 - Baza podataka je snimljena pomoću 10 glumaca koji su izražavali emocije na osnovu zadatih scenarija, a ima i improvizacija
 - Sadrži audio, video i audio transkripte (tekst)
 - Ideja: pomoću ovih podataka formirati i kombinovati tri različita klasifikatora radi maksimizacije postignute tačnosti klasifikacije emocija
 - Mogla bi se isprobati neka varijanta sa polu-nadgledanim obučavanjem, npr. co-training bi bio dobar kandidat s obzirom da ima više nezavisnih izvora podataka
 - Moglo bi se pokušati da se spoji ova baza sa drugim slobodnim emotivnim audio bazama kako bi se maksimizovala tačnost
 - Obratiti se Jeleni (slivkaje@uns.ac.rs) za podatke



Izrada chatbota

- Agent koji interaguje sa korisnicima na odredjenu temu putem rečenica prirodnog jezika
- Studenti su napravili chatbot koji sa korisnikom komunicira u stilu govora Yode iz filma Star Wars
- Obučavajući skup je bio skup transkripata dijaloga iz filmova
- Podaci su prečišćeni i rečenice su analizirane korišćenjem NLP tehnika
- Za obučavanje je korišćena RNN

Još neke ideje – predviđanje cena akcija

- Predviđanja budućih cena akcija Apple na osnovu istorijskih cena
 - Poređenje postojećih matematičkih modela sa tehnikama istraživanja podataka
- Na osnovu sadržaja novinskih članaka (teme i sentimenta) predvideti promenu berzanskih vrednosti
 - Pokazano je da postoji korelacija između raspoloženja twitter korisnika i berzanskog indeksa
 - Pokazano je da postoji korelacija između objavljenih novinskih članaka i vrednosti akcija



Predviđanje cena krypto valuta

- Dejan Kuzmanović i Milica Milutinović
- Sami prikupljali podatke
- Predikcija je vršena na osnovu:
 - Istorijskih informacija (Coinmarketcap API i Poloniex API – cena na početku intervala, cena na kraju intervala, najviša/najniža cena,...)
 - Informacija o blockchainu (*block size, cost per transaction, difficulty, hash rate, transactions per block,...*)
 - Sentimentalne analize vesti i tvitova (Twitter REST API – dosta posla oko pretprocesiranja teksta. Leksički zasnovana analiza sentimenata)

Predikcija cene nekretnine

- Ivan Radosavljević, Aleksandra Mitrović, Mladen Vidović
- Sami sakupili podatke sa veb stranice za oglašavanje nekretnina nekretnine.rs
- Bazirana na:
 - Tehničkim specifikacijama
 - Površina, datum objave oglasa, broj soba, kupatila,...
 - Slikama
 - Neuronska mreža prepoznaje postoje li odeđeni objekti (klima uređaj, bojler, veš mašina, sudo-mašina,...) – definisanje opremljenosti nekretnine
 - Susedstvo (geografska lokacija)
 - Na osnovu koordinate lokacije detektovano je postojanje okolnih objekata koji bi mogli uticati na kvalitet susedstva: škole, turistički obekti, industrijske zone, mesta za odmor i opuštanje,...
- Primenjeno je više regresionih modela, urađena opsežna eksplorativna analiza i analiza grešaka modela

Predikcija broja oružanih sukoba

- Potrebu za ovakvim sistemom imaju mirovne organizacije (npr. Ujedinjene nacije)
- Sistem bi omogućio usmeravanje resursa na sprečavanje konflikata i saniranje posledica
- Na osnovu istorijskih podataka o konfliktima
 - ACLED (*The Armed Conflict Location & Event Data Project*) – podaci o bitkama, protestima, pobunama,...
- Na osnovu socioekonomskih pokazatelja:
 - Broj stanovnika, BDP, površina neobrađivog zemljišta, izvoz robe, etička podeljenost, obrazovanje,...
- Pozitivni aspekti:
 - Agregacija više izvora podataka
 - Pažljiva eksplorativna analiza skupa podataka i rezultujući data wrangling, dubinska analiza domena