



HANDWRITTEN TEXT RECOGNITION (HTR)

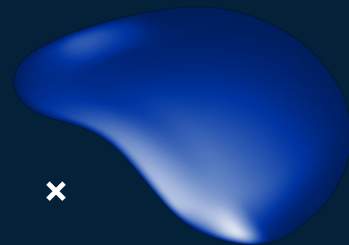
BOSANSKI JEZIK - ŠTAMPANA SLOVA

Elektrotehnički fakultet Univerziteta u Sarajevu
Vještačka inteligencija
Studenti: Ivona Jozić i Ismar Muslić

Juni 2025. god.



SADRŽAJ



01 **PREGLED STANJA U
OBLASTI**

Historijski pregled, CNN,
RNN, LLM

02 **DATASET**

Kreiranje, preprocesiranje
dataset-a

03 **MODEL**

Opis, treniranje

04 **REZULTATI**

Metrike, dobiveni rezultati,
usporedba sa LLM

05 **ZAKLJUČAK**



UVOD

- Sistem za prepoznavanje rukom pisanog teksta
- Vještačka inteligencija, računarska vizija i obrada prirodnog jezika
- Ogromne varijabilnosti u načinu pisanja

KORACI



DATASET



MODEL



REZULTATI

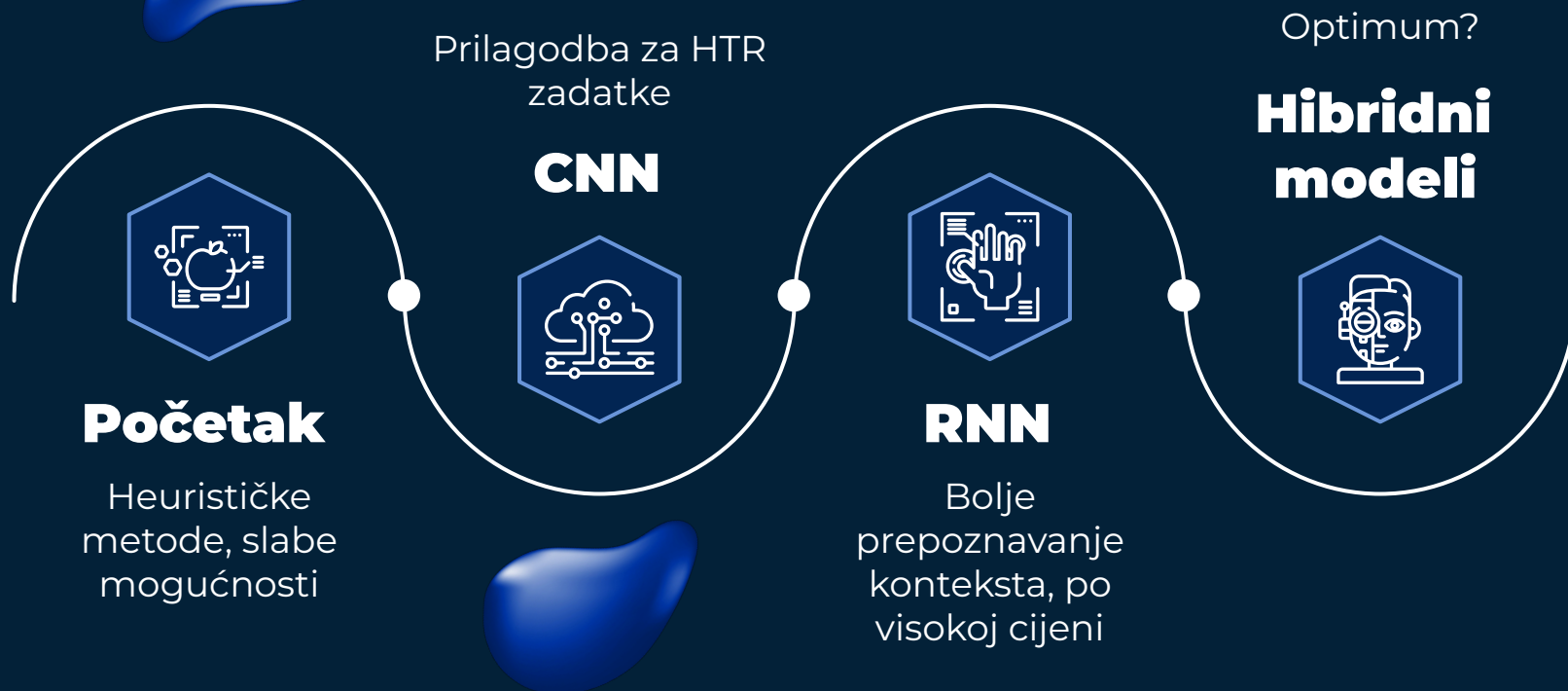




01

PREGLED STANJA U OBLASTI

TIMELINE RAZVOJA*

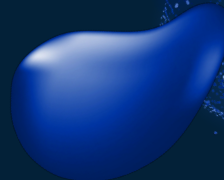


02

DATASET



+



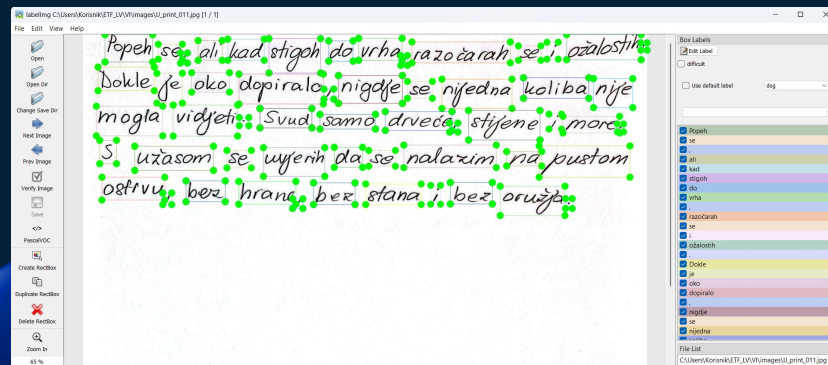
x

x

KREIRANJE DATASET-a

- Dataset kreiran tokom kursa
- Pronaći tekstove
- Rukom napisati tekstove u 3 stila
- Skenirati i označiti svaki napisani tekst
- Izdvojeni tekstovi pisani štampanim slovima - 156 fotografija

Popeh se, ali kad stigoh do vrha, razočarah se i ožalostih.
Dokle je oko dopiralo, nigdje se nijedna koliba nije
mogla vidjeti. Svud samo drveće, stijene i more.
S užasom se uviđah da se nalazim na pustom
ostrvu, bez hrane, bez stana i bez oružja.

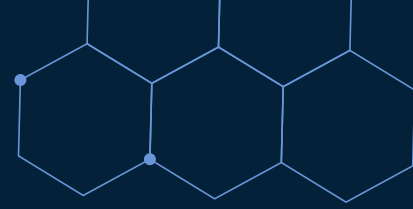


PREPROCESIRANJE DATASET-a

- HDF5 format
- Izvlačenje pojedinačnih riječi iz tekstova
- Izvlačenje linija teksta

čovječuljak

Obavještavaju se studenti da će se naredno



IZVLAČENJE RIJEČI IZ TEKSTA

- Prolazi se kroz sve .xml i kreira word_labels.txt
- Na osnovu njega će se vršiti izrezivanje slika
- Folder labels_w i u njemu slike riječi
- Izrezano 11 529 riječi

```
LB_print_005.jpg|0|3307|4676|81|86|147|160|"  
LB_print_005.jpg|1|3307|4676|149|118|331|248|AKO  
LB_print_005.jpg|2|3307|4676|385|128|513|230|TE  
LB_print_005.jpg|3|3307|4676|569|122|905|246|PUSTIM  
LB_print_005.jpg|4|3307|4676|927|180|973|260|,  
LB_print_005.jpg|5|3307|4676|1001|88|1321|234|HOĆEŠ  
LB_print_005.jpg|6|3307|4676|1363|106|1445|218|LI  
LB_print_005.jpg|7|3307|4676|1487|92|1841|226|DRŽATI  
LB_print_005.jpg|8|3307|4676|1891|106|2177|246|JEZIK  
LB_print_005.jpg|9|3307|4676|2203|112|2331|214|ZA
```

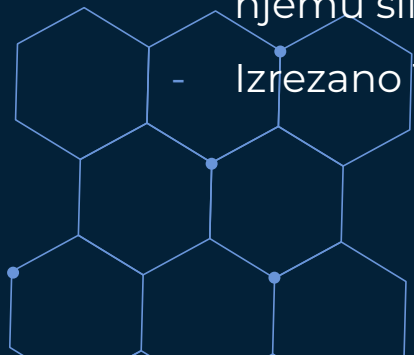
filename|img_width|img_height|word_idx|xmin|ymin|xmax|ymax|word

11

Ako

te

pustim



IZVLAČENJE LINIJA IZ TEKSTA

- Prolazi se kroz sve .xml i kreira line_labels.txt
- Na osnovu njega će se vršiti izrezivanje linija
- Folder labels_l i u njemu linije teksta
- Izrezano 1 497 linija

```
LB_print_005.jpg|0|3307|4676|81|84|2927|160|" "  
LB_print_005.jpg|1|3307|4676|149|88|2875|248|AKO TE PUSTIM HOĆEŠ LI DRŽATI JEZIK ZA ZUBIMA ?  
LB_print_005.jpg|2|3307|4676|927|180|973|260|,  
LB_print_005.jpg|3|3307|4676|81|290|2279|440|NISAM SE POTRUDIO DA MU ODGOVORIM .  
LB_print_005.jpg|4|3307|4676|79|510|131|582|" "  
LB_print_005.jpg|5|3307|4676|135|524|2873|698|HOLDENE AKO TE PUSTIM , HOĆEŠ LI DRŽATI JEZIK ZA  
LB_print_005.jpg|6|3307|4676|521|632|555|694|,  
LB_print_005.jpg|7|3307|4676|141|692|1705|838|ZUBIMA ? REKAO JE PONOVO .  
LB_print_005.jpg|8|3307|4676|587|686|631|748|" "  
LB_print_005.jpg|9|3307|4676|655|786|699|854|,  
LB_print_005.jpg|10|3307|4676|85|892|507|968|" "  
LB_print_005.jpg|11|3307|4676|163|918|429|1038|HOĆU  
LB_print_005.jpg|12|3307|4676|431|992|489|1038|.  
KH_print_000.jpg|0|1179|1692|36|24|1088|130|PRILIKOM RJEŠAVANJA PROBLEMA LINEARNOG PROGRAMIRANJA ,  
KH_print_000.jpg|1|1179|1692|40|119|997|210|ČESTO JE , RADI LAKŠEG RAČUNANJA , POŽELJNO VRŠITI
```

**filename|img_width|img_height|line_idx|line_x_min|line_y_min|
line_x_max|line_y_max|text**

zubima?', rekao je ponovo.

Rekao sam da znam. Postao sam dobro upućen u

Naziv skupa metoda, koje imaju za cilj nalaženje najboljih (optimalnih) rješenja složenih problema, je relativno novi

"...da li dođete ispitati zubima?"

?

.

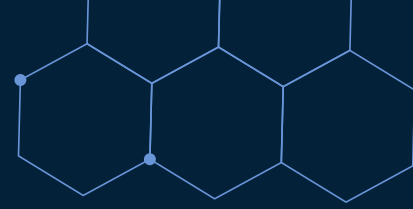
||

IZVLAČENJE LINIJA IZ TEKSTA - MODIFIKOVANO

- Prolazi se kroz sve .xml i kreira line_labels_m.txt
- Na osnovu njega će se vršiti izrezivanje linija
- x koordinata se posmatra
- Folder labels_l_m i u njemu linije teksta
- Izrezano 1 359 linija

```
LB_print_005.jpg|0|3307|4676|81|84|2927|260|" AKO TE PUSTIM, HOĆEŠ LI DRŽATI JEZIK ZA ZUBIMA? "  
LB_print_005.jpg|1|3307|4676|81|290|2279|440|NISAM SE POTRUDIO DA MU ODGOVORIM.  
LB_print_005.jpg|2|3307|4676|79|510|2873|698|" HOLDENE, AKO TE PUSTIM, HOĆEŠ LI DRŽATI JEZIK ZA  
LB_print_005.jpg|3|3307|4676|141|686|1705|854|ZUBIMA? ", REKAO JE PONOVO.  
LB_print_005.jpg|4|3307|4676|85|892|507|1038|" HOĆU. "  
KH_print_000.jpg|0|1179|1692|36|24|1088|130|PRILIKOM RJEŠAVANJA PROBLEMA LINEARNOG PROGRAMIRANJA,  
KH_print_000.jpg|1|1179|1692|40|119|997|210|ČESTO JE, RADI LAKŠEG RAČUNANJA, POŽELJNO VRŠITI  
KH_print_000.jpg|2|1179|1692|43|196|1091|294|SKALIRANJA OGRANIČENJA. MEĐUTIM, KAD GOD TO RADIMO,  
KH_print_000.jpg|3|1179|1692|44|284|1029|381|POTREBNO JE VODITI RAČUNA O MJERNIM JEDINICAMA.  
KH_print_000.jpg|4|1179|1692|44|359|1107|464|NAIME, BEZ OBZIRA ŠTO ĆE RJEŠENJE UVIJEK BITI TAČNO
```

**filename|img_width|img_height|line_idx|line_x_min|line_y_min|
line_x_max|line_y_max|text**

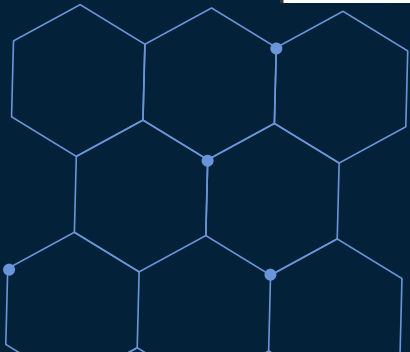


"Ako te pustim, hoćeš li držati jezik za zubima?"

Naziv skupa metoda, koje imaju za cilj nalaženje najboljih

zasnovan na sposobnosti kompetitivnog vezivanja hemijske materije

4. 8. 2025.



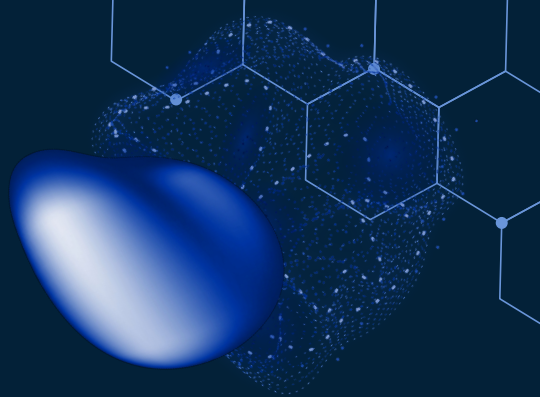
PODJELA NA TRENING I TEST SKUP PODATAKA

- Koristi `train_test_split` iz `sklearn.model_selection`
- Na osnovu kreiranih `word.txt` i `line.txt` u toku izrezivanja riječi i linija vrši podjelu i spremanje odgovarajućih labels u tri različita `.txt` - `trainset`, `testset` i `validset`
- Za treniranje se koristi **70%** podataka iz dataset-a
- Za validaciju se koristi **20%** podataka
- Za testiranje se koristi **10%** podataka

RIJEČI

trening (8113), validacija (2319), test (1160)

×



×



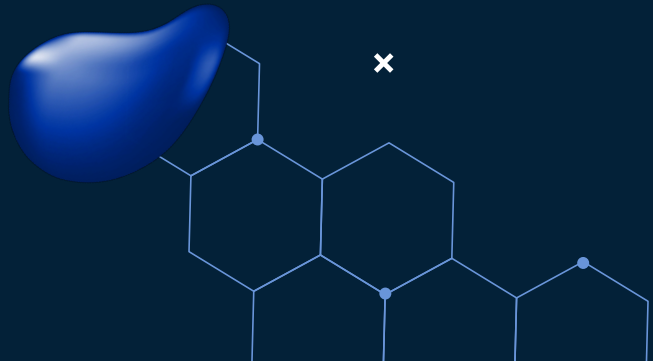
LINIJE

trening (1047), validacija (300), test (150)

LINIJE_M

trening (951), validacija (272), test (136)

×

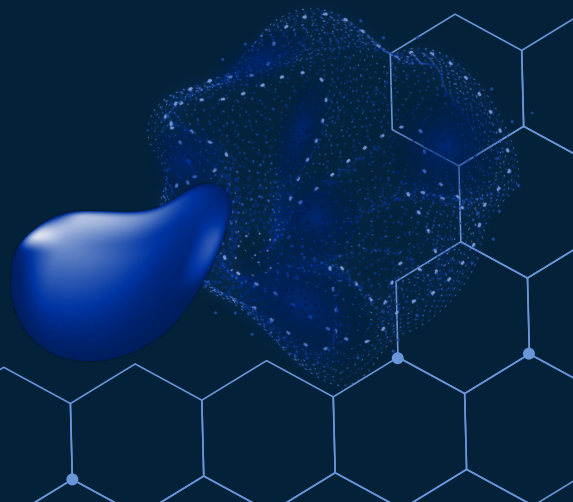


03

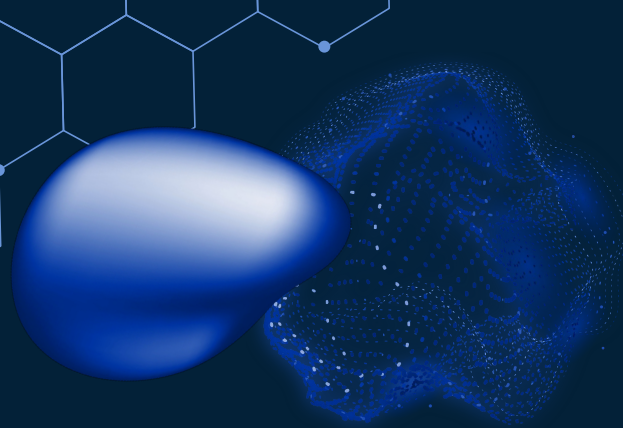
MODEL



+



x

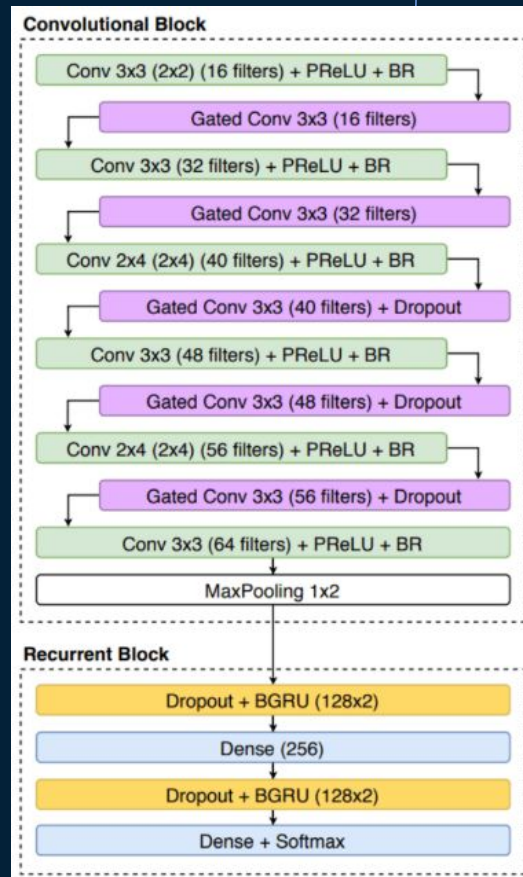


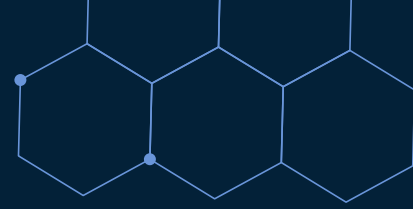
x



ARHITEKTURA NEURALNE MREŽE

- FLORe mreža
- Ulaz u mrežu je slika
- 6 konvolucijskih blokova

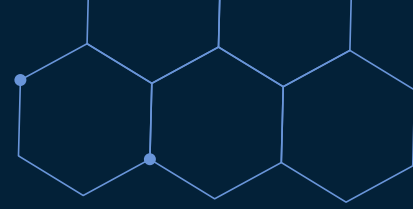




KONFIGURACIJA

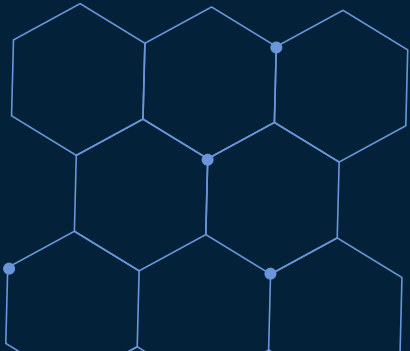
- Google Drive mount
- TensorFlow 2.x
- Definisati izvor podataka
- Broj epoha (1000 uz EarlyStopping)
- Batch size (64 i 16)
- Podešen set podataka - ASCII + afrikati
- DataGenerator, HTRModel klase





TRENIRANJE MODELA

- Prati se ukupno trajanje treniranja
- Trajanje pojedinačnih epoha
- Podaci o treniranju se spašavaju na GoogleDrive
- EarlyStopping
- Resursi na Google Colab - u

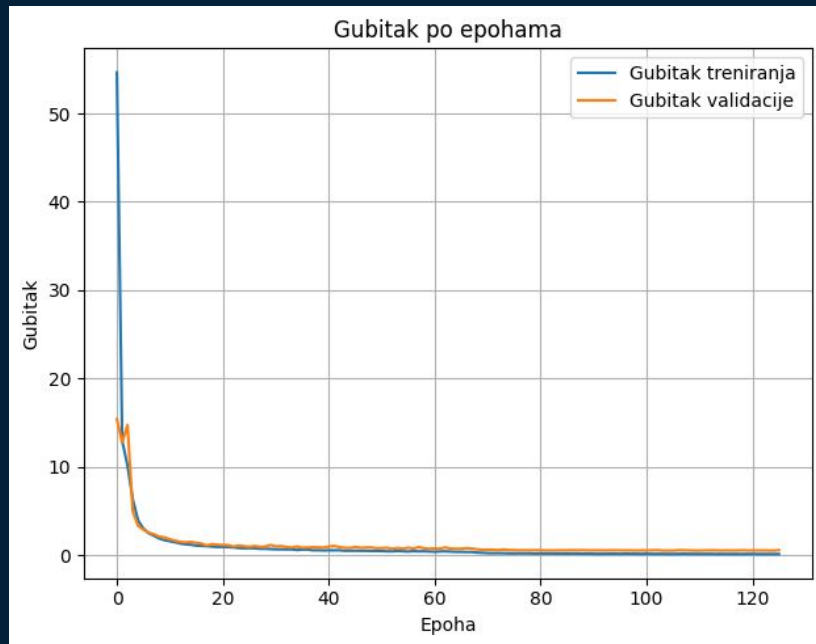


TRENIRANJE NA RIJEČIMA

Total time: 3:54:22.086345
Time per epoch: 0:01:51.603860
Time per item: 0:00:00.010698

Total epochs: 126
Best epoch 106

Training loss: 0.11550388
Validation loss: 0.50083512

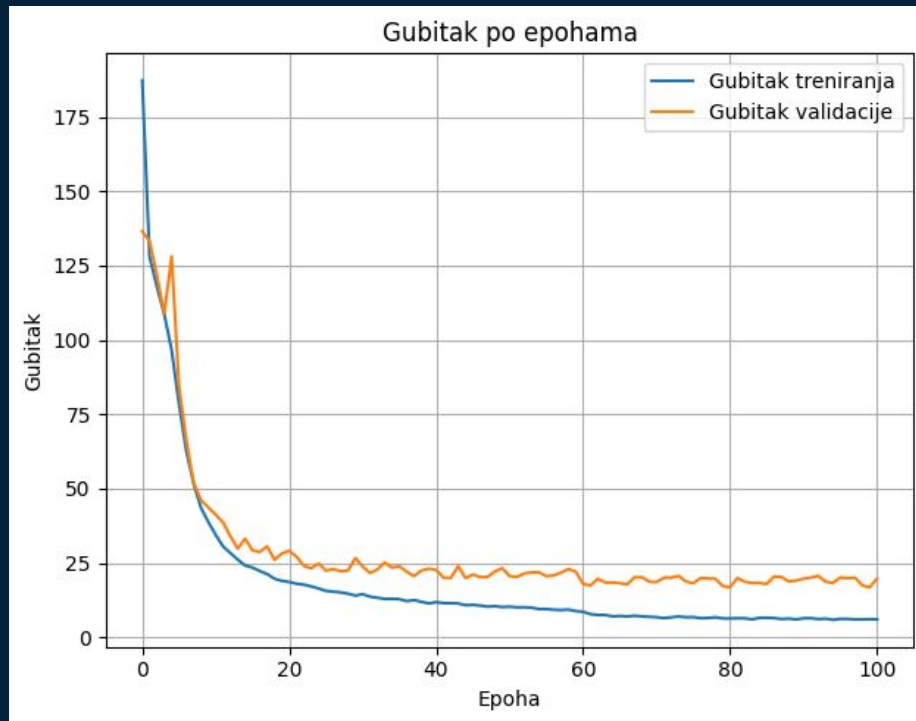


TRENIRANJE NA LINIJAMA

Total time: 0:32:03.886557
Time per epoch: 0:00:19.048382
Time per item: 0:00:00.014162

Total epochs: 101
Best epoch: 81

Training loss: 6.34184980
Validation loss: 16.79165840

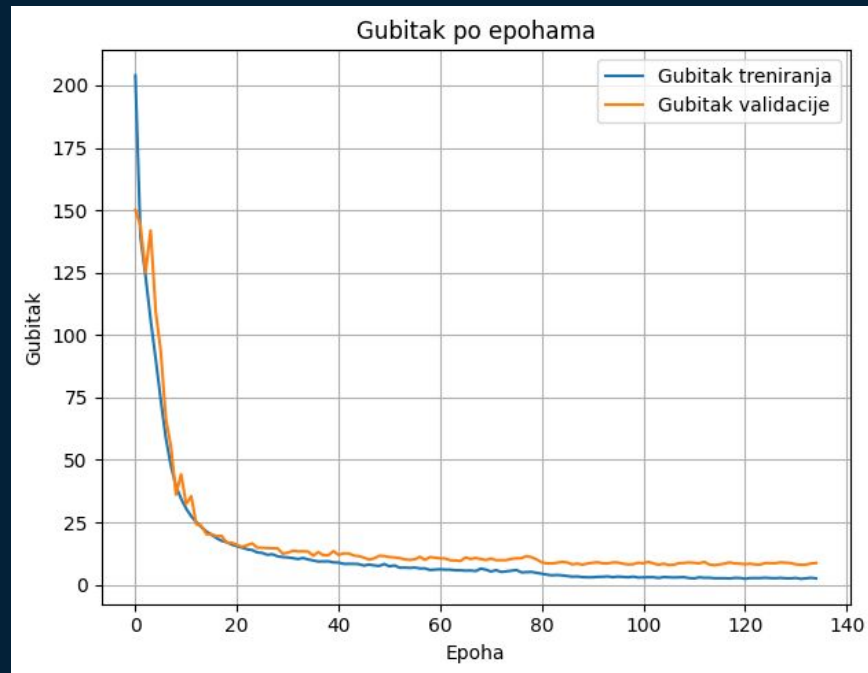


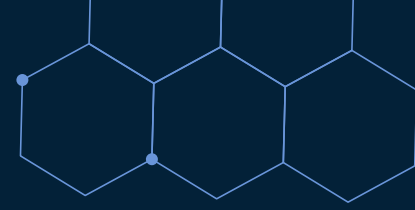
TRENIRANJE NA LINIJAMA - MODIFICIRANIM

Total time: 0:42:03.257401
Time per epoch: 0:00:18.690796
Time per item: 0:00:00.015283

Total epochs: 135
Best epoch 115

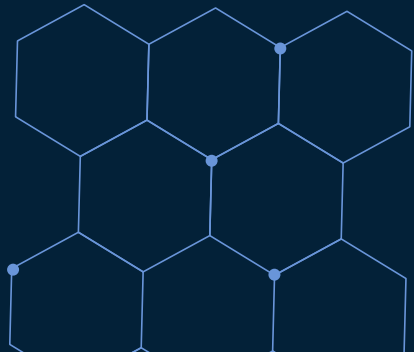
Training loss: 2.61487579
Validation loss: 7.84266376





TESTIRANJE MODELA I METRIKE

- Preprocess metoda
- Prikaz nekoliko prvih rezultata
- Metrike WER, CER, SER - ocr_metrics metoda

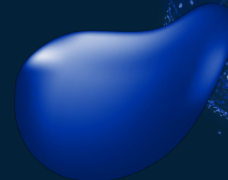
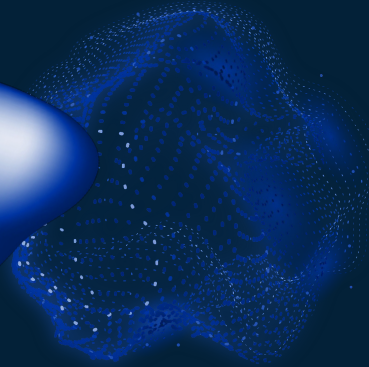
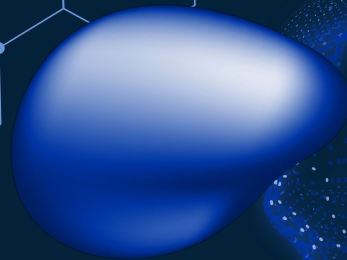


04

ANALIZA REZULTATA

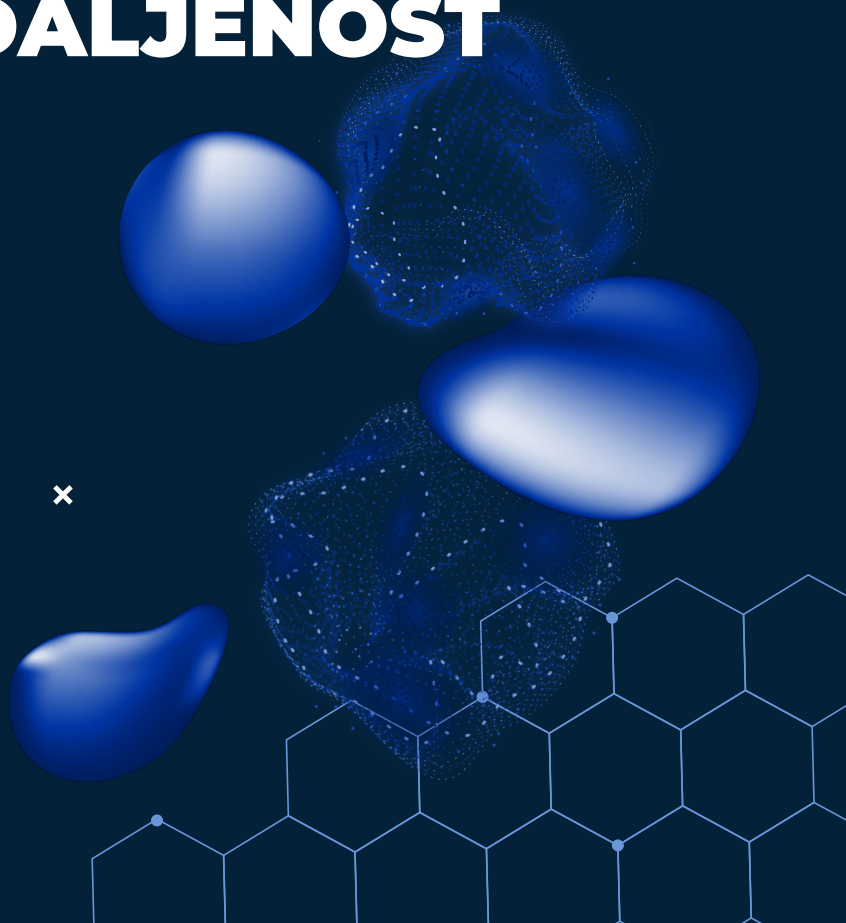


+



LEVENSHTEIN UDALJENOST

- ◆ Metrike za analizu se zasnivaju na *Levenshtein* udaljenosti
- ◆ Mjera zasnovana na tri operacije: *izmjena*, *brisanje*, *umetanje*
- ◆ “Stol” \longrightarrow “Stop” : distanca 1



CER, WER, SER

- ◆ CER: “*Character Error Rate*”, greške na nivou individualnih znakova
- ◆ WER: “*Word Error Rate*”, greške na nivou riječi
- ◆ SER: “*Sequence Error Rate*”, greške na nivou sekvence
- ◆ SER najstrožija

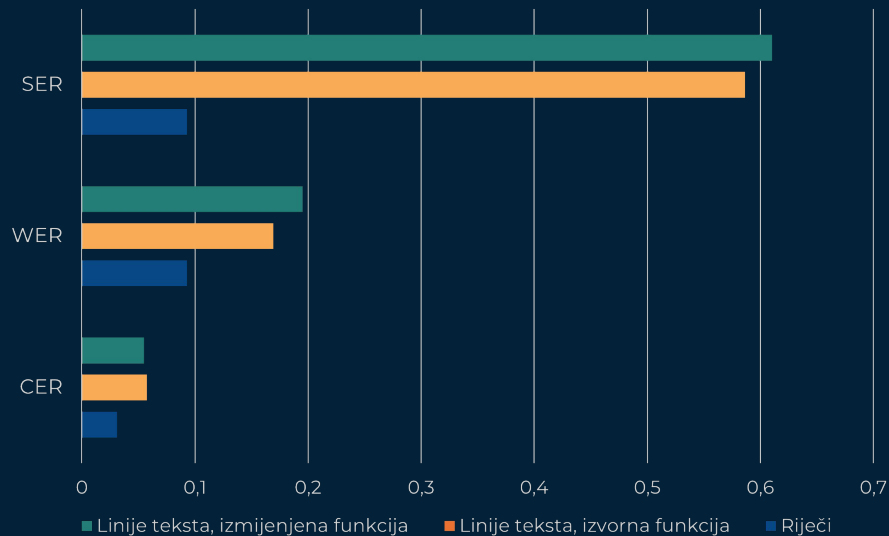
$$CER = \frac{S + D + I}{N} \times 100\%$$

$$WER = \frac{S + D + I}{N} \times 100\%$$

$$SER = \frac{N_{err}}{N_{total}} \times 100\%$$

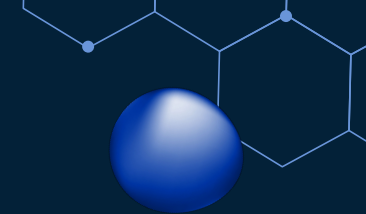
METRIKE

- ◆ CER u skladu s očekivanjima
- ◆ SER bolji kod funkcije koja griješi?
- ◆ WER isti kao SER za riječi?



Rezultati metričke analize

GREŠKE MODELA



(za snižavanje holesterola u krvi). Ako

Ispravna labela: (ZA SNIŽAVANJE HOLESTEROLA U KRVU) . AKO

Predikcija modela: LZA SNIŽAVANJE HOLESTEROLA U KRVU) . AKO

IT inženjering je oblasť koja se bavi razvojem, održavanjem

Ispravna labela: IT INŽENJERING JE OBLAST KOJA SE BAVI RAZVOJEM, ODRŽAVANJEM

Predikcija modela: 1? INŽENJERING JE OBLASĆ KOJA SE BAVI RAZVOJEM, ODIŽAVANJEM



DA LI LLM VIDI GREŠKE?



- ◆ GPT-4o, Claude Sonnet 4, Gemini 2.5 - flash
- ◆ LLM-ovi trenirani većinom na engleskom jeziku
- ◆ Bolje sposobnosti razumijevanja i kontekstualizacije
- ◆ Poslane iste slike linija teksta
- ◆ Standardiziran upit
- ◆ Koliko precizni će biti rezultati?

"I am uploading an image that contains handwritten text. Please scan the image, recognize the handwritten text, and return the result as clean, digitalized plain text.

Do not include any formatting, background information, or interpretation—only transcribe the visible handwriting as accurately as possible.

Preserve punctuation characters if any are visible in the handwriting

If any parts of the handwriting are unclear or illegible, please mark them with [unreadable]."

Tekst upita



GPT-4o

(za snižavanje holesterola u krvi). Ako

I am uploading an image that contains handwritten text. Please scan the image, recognize the handwritten text, and return the result as clean, digitalized plain text.

Do not include any formatting, background information, or interpretation—only transcribe the visible handwriting as accurately as possible.

Preserve punctuation characters if any are visible in the handwriting

If any parts of the handwriting are unclear or illegible, please mark them with [unreadable].

(za snižavanje holesterola u krvi). Ako



IT inženjering je oblast koja se bavi razvojem, održavanjem

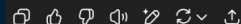
I am uploading an image that contains handwritten text. Please scan the image, recognize the handwritten text, and return the result as clean, digitalized plain text.

Do not include any formatting, background information, or interpretation—only transcribe the visible handwriting as accurately as possible.

Preserve punctuation characters if any are visible in the handwriting

If any parts of the handwriting are unclear or illegible, please mark them with [unreadable].

IT inženjering je oblast koja se bavi razvojem, održavanjem



CLAUDE SONNET 4



- I** I am uploading an image that contains handwritten text. Please scan the image, recognize the handwritten text, and return the result as clean, digitalized plain text.
- * Do not include any formatting, background information, or interpretation—only transcribe the visible handwriting as accurately as possible.
 - * Preserve punctuation characters if any are visible in the handwriting
 - * If any parts of the handwriting are unclear or illegible, please mark them with [unreadable].

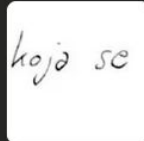
Looking at the handwritten text in the image, I can transcribe:

za snižavanje holesterola u krvi. Ako



📄 👍 🗨️ [Retry](#) ▾

Claude can make mistakes. Please double-check responses.



- I** I am uploading an image that contains handwritten text. Please scan the image, recognize the handwritten text, and return the result as clean, digitalized plain text.
- * Do not include any formatting, background information, or interpretation—only transcribe the visible handwriting as accurately as possible.
 - * Preserve punctuation characters if any are visible in the handwriting
 - * If any parts of the handwriting are unclear or illegible, please mark them with [unreadable].

Looking at the handwritten text in the image, I can transcribe:

IT inženjering je oblik koji se bavi razvojem, održavanjem



📄 👍 🗨️ [Retry](#) ▾

Claude can make mistakes. Please double-check responses.

GEMINI 2.5-FLASH



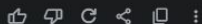
(za snižavanje holesterola u krvi). Ako

I am uploading an image that contains handwritten text.
Please scan the image, recognize the handwritten text,
and return the result as clean, digitalized plain text....



Analysis ▼

(za snižavanje holesterola u krvi). Ako



IT inženjering je oblas koja se bavi razvojem, održavanjem

I am uploading an image that contains handwritten text.
Please scan the image, recognize the handwritten text,
and return the result as clean, digitalized plain text....



Analysis ▼

IT inženjering je oblast koja se bavi razvojem, održavanjem





05

ZAKLJUČAK



ZAKLJUČNE TAČKE



- ◆ Uspješno treniran HTR model, uz zadovoljavajuće rezultate
- ◆ Najbolje metrike (CER 3.12%, WER/SER 9.3%) postignute treniranjem na pojedinačnim riječima.
- ◆ Poboljšanja bi uključivala veći i kvalitetniji dataset te treniranje na snažnijem lokalnom hardveru.
- ◆ LLM-ovi pokazuju potencijal za HTR; ograničeni su nedostatkom podataka na bosanskom jeziku.
- ◆ Hibridne arhitekture (CNN+RNN) ostaju najučinkovitije rješenje za složene HTR zadatke u budućnosti.



**HVALA NA
PAŽNJI**

