

Music Emotion Classification

– LSTM i Transformer –

Julijana Jevtić Jelena Milošević

Matematički fakultet, Univerzitet u Beogradu

Septembar, 2025

1 Uvod

2 Modeli

- Long Short-Term Memory (LSTM)
- Transformer
- Loss and Accuracy
- Metrički rezultati modela
- F1 skor po klasama
- Normalizovana matrica konfuzije
- ROC krive po klasama
- Precision-recall krive

3 Zaključak

- Klasifikacija emocija iz muzičkih tekstova
- Cilj – prepoznavanje emocija (sreća, tuga, ljubav, ljutnja)
- Primena u muzičkim preporukama i personalizaciji
- Zadatak podrazumeva **analizu teksta** i **modelovanje sekvenci**

LSTM - Long Short-Term Memory

Sekvencijalni model za dugoročne zavisnosti.

Transformer

Model zasnovan na pažnji (attention), efikasniji u NLP zadacima.

Long Short-Term Memory (LSTM)

- Vrsta RNN-a koja rešava problem **nestajanja gradijenata**.
- Koristi **memorijske ćelije** i **gate mehanizme** (input, forget, output) za kontrolu informacija.
- Obrada sekvenci:
 - ① Zaboravljanje irelevantnih podataka.
 - ② Dodavanje novih informacija.
 - ③ Ažuriranje i prosleđivanje značajnih podataka.
- **Primene:** NLP (analiza, prevođenje, generisanje teksta), prepoznavanje govora, vremenske serije, biomedicina, računarska vizija, robotika, optimizacija resursa...
- **Prednosti:** stabilno učenje, dobar za dugoročne zavisnosti.
- **Ograničenja:** složeniji od RNN-a, osetljiv na inicijalizaciju.

Transformer

- Arhitektura zasnovana na **mehanizmu pažnje** (self-attention).
- Obrada celih sekvenci **paralelno**, za razliku od RNN-a.
- Ključne komponente:
 - 1 **Self-attention** za hvatanje odnosa između svih reči.
 - 2 **Multi-head pažnja** za učenje različitih tipova zavisnosti.
 - 3 **Pozicionalno kodiranje** za očuvanje redosleda.
 - 4 **Feedforward mreže i rezidualne veze.**
- **Primene:** mašinsko prevođenje, sažimanje teksta, generisanje teksta, odgovaranje na pitanja, obrada slika i govora.
- **Prednosti:** skalabilnost, hvatanje dugoročnih zavisnosti, osnova savremenih modela (BERT, GPT, T5...).
- **Ograničenja:** visoka složenost, velika potrošnja memorije i resursa.

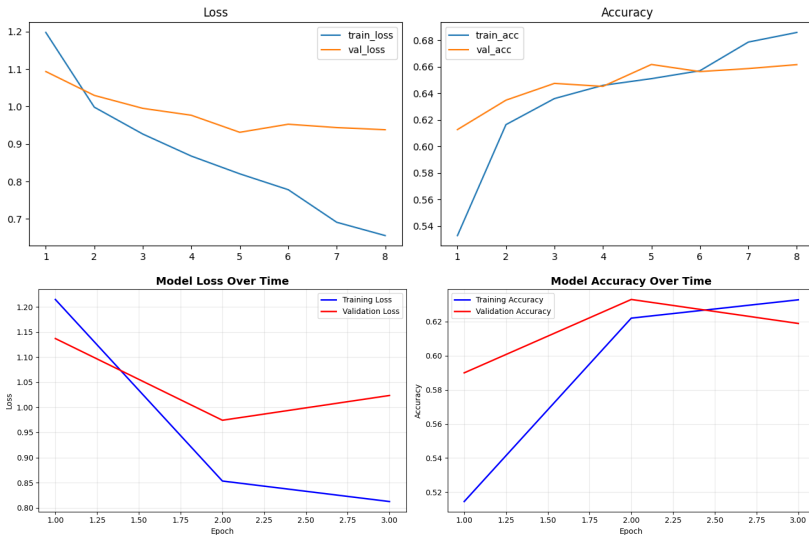
Rezultati primene modela – Hiperparametri

U ovom poglavlju prikazani su rezultati evaluacije LSTM i Transformer modela nad test skupom podataka

LSTM		Transformer	
Veličina rečnika	20 000	Veličina rečnika	10 000
Maks. dužina sekvence	128	Maks. dužina sekvence	384
Dim. ugnežđavanja	100	Dim. ugnežđavanja	48
LSTM jedinice	64	Attention heads	4
Gusti sloj (neurona)	32	Dim. feed-forward sloja	192
Dropout	0.3	Dropout	0.2
Batch veličina	64	Batch veličina	16
Broj epoha	8	Broj epoha	6
Random seed	42	Patience	1
		Learning rate	$2 \cdot 10^{-4}$

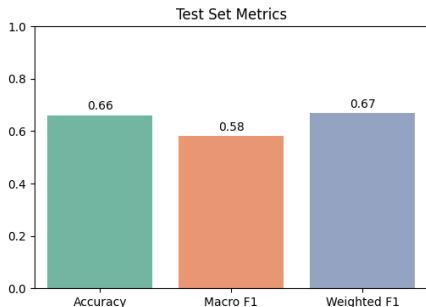
Tabela: Uporedni prikaz hiperparametara LSTM i Transformer modela.

Tok funkcije gubitka i tačnosti tokom epoha

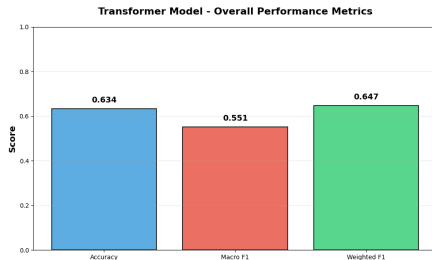


Slika: LSTM krive (gore) i Transformer krive (dole)

Metrički rezultati modela



Slika: LSTM metrike

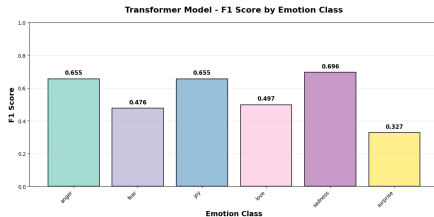


Slika: Transformer metrike

F1 skor po klasama

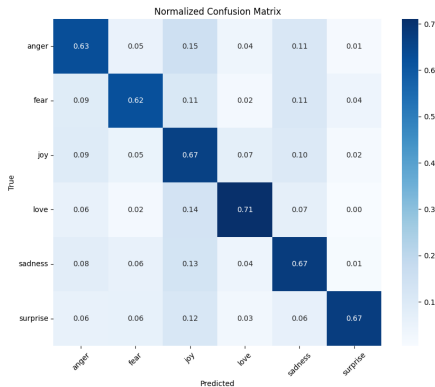


Slika: LSTM F1 Score

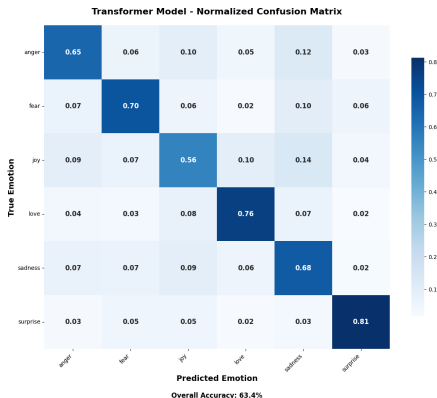


Slika: Transformer F1 Score

Normalizovana matrica konfuzije



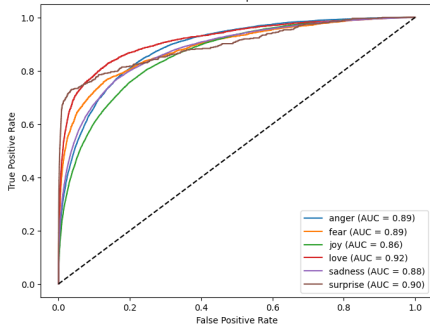
Slika: LSTM matrica



Slika: Transformer matrica

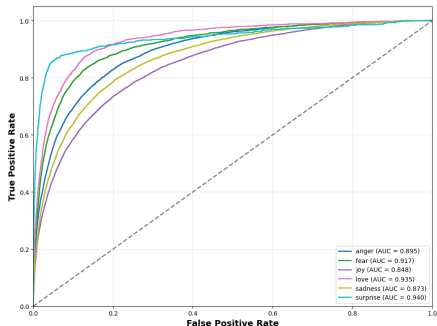
ROC krive po klasama

ROC Curves per Class



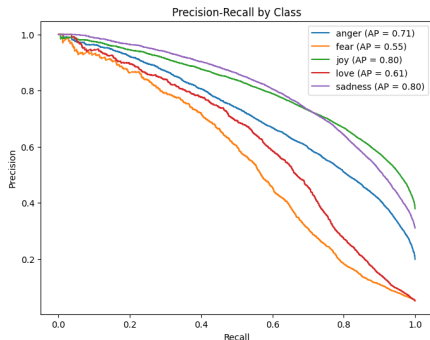
Slika: LSTM kriva

Transformer Model - ROC Curves per Emotion Class

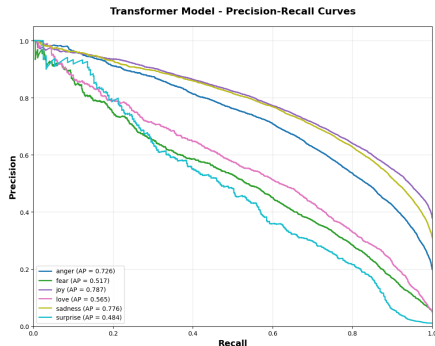


Slika: Transformer kriva

Precision-recall krive



Slika: LSTM kriva



Slika: Transformer kriva

Primeri predviđanja na osnovu teksta

Text snippet	True	Pred	Prob
"i told you that i loved you and i meant it..."	joy	love	0.798
"merry christmas have a very very merry..."	joy	joy	0.999
"this ending is all but an ending he said..."	anger	sadness	0.727
"i catch a vibe when i am with you lets..."	joy	joy	0.927
"so many memories and so many miles the..."	joy	sadness	0.614

Tabela: Primeri predikcija LSTM modela za emocije u pesmama (skraćeni tekstovi).

Primeri predviđanja na osnovu teksta (Transformer)

Text snippet	True	Pred	Prob
".she got the red wine in her glass she is..."	joy	joy	0.605
"and god said let there be light kings..."	sadness	sadness	0.798
"got to take the shame from my back it..."	sadness	fear	0.348
"merry christmas have a very very merry..."	joy	joy	0.971
"i catch a vibe when i am with you lets..."	joy	sadness	0.599

Tabela: Primeri predikcija Transformer modela za emocije u pesmama (skraćeni tekstovi).

- Poređenje LSTM i Transformer modela dalo je **vrlo slične rezultate**.
- Uprkos **dominaciji** Transformer arhitektura u savremenim NLP zadacima, LSTM modeli su i dalje **relevantni i konkurentni** modeli.