

# Predviđanje nivoa depresije kod ljudi na osnovu audio zapisa

Analiza akustičkih karakteristika govora za detekciju nivoa depresije

Veselin Roganović, SV 36/2022 | Fakultet tehničkih nauka, Novi Sad



## Uvod

Depresija predstavlja jedan od najvećih izazova modernog društva. Pravovremena detekcija je ključna, ali često teška čak i za stručnjake. Ovaj projekat istražuje mogućnost automatske procene nivoa depresije na osnovu **akustičkih karakteristika govora**, bez analize sadržaja izgovorenih reči.

## Skup podataka

- Korišćen je **DAIC-WOZ** skup podataka koji sadrži:
- Audio zapise kliničkih intervjua.
  - PHQ-8** skorove (0-24) kao ciljnu vrednost.
  - Ukupno 140 uzoraka (80% trening, 10% val, 10% test).

## Predobrada

- Resampling na 16kHz i normalizacija amplitude.
- Ekstrakcija MFCC koeficijenata, energije i pitch-a.

## Metodologija

Primenjena su dva različita pristupa:

### 1. SVR (Support Vector Regression):

- Ulaz: Statistike MFCC-a (prosek, std, percentili).
- Optimizacija: RBF kernel, Grid-Search za  $C$  i  $\epsilon$ .

### 2. CNN + LSTM:

- Ulaz: MFCC matrice (vremenski prozori).
- Arhitektura: Konvolucioni slojevi za prostorna obeležja praćeni LSTM slojevima za sekvence.

## Arhitektura CNN+LSTM

- CNN:** 2D slojevi sa BatchNormalization i Dropout-om.
- LSTM:** 2 sloja sa 64 skrivene jedinice.
- Gubitak:** Kombinovani MAE i Pearson loss.

## Rezultati

Model	MAE	RMSE	Prsn
SVR	4.65	5.32	<b>0.57</b>
CNN+LSTM	<b>3.99</b>	<b>5.03</b>	0.45

CNN+LSTM model postiže preciznija predviđanja (niži MAE), dok SVR pokazuje bolju linearnu korelaciju.

## Zaključak

Rezultati potvrđuju da akustička obeležja nose značajne informacije o mentalnom stanju govornika. Kombinacija konvolucionih i rekurentnih mreža daje obećavajuće rezultate.

## Korišćeni alati i Literatura

Python, PyTorch, Librosa, Scikit-Learn  
1. Alhanai et al., "Detecting Depression with Audio/Text Sequence Modeling", 2018.