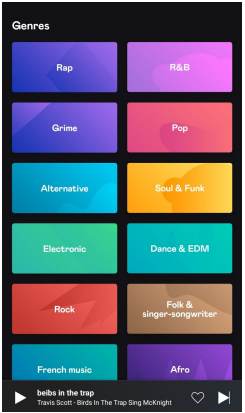
Klasifikacija muzičkih žanrova

Natalija Asanović
Lucija Miličić

Matematički fakultet
September 9, 2022

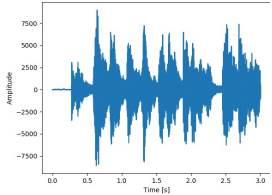
Motivacija

- Organizovanje pesama u različite kategorije
- Automatizacija muzičke klasifikacije

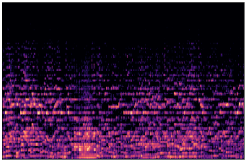


Vizuelizacija audio zapisa

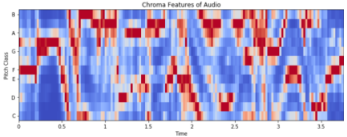
- grafik amplitude zvučnih talasa



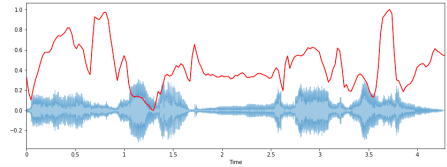
- spektrogram



- chroma feature



- spectral rolloff



Spektrogram

- vizualna reprezentacija spektra frekvencija zvuka ili drugih signala kako variraju kroz vreme
- može se dobiti kratkoročnom Furijeovom transformacijom (STFT)

```
D = librosa.stft(y)  # STFT of y
S_db = librosa.amplitude_to_db(np.abs(D), ref=np.max)

plt.figure()
librosa.display.specshow(S_db)
plt.colorbar()
```

Dataset

GTZAN dataset:

- 10 žanrova, po 100 pesama
- odgovarajući spektrogrami
- akustičke karakteristike (.csv)

Neuronske mreže

Potpuno povezana

Layer (type)	Output Shape	Param #
dense (Dense)	(None, 256)	15104
dropout (Dropout)	(None, 256)	0
dense_1 (Dense)	(None, 128)	32896
dropout_1 (Dropout)	(None, 128)	0
dense_2 (Dense)	(None, 64)	8256
dropout_2 (Dropout)	(None, 64)	0
dense_3 (Dense)	(None, 10)	650

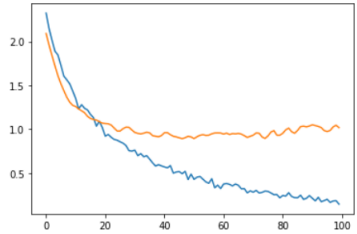
Total params: 56,906
Trainable params: 56,906
Non-trainable params: 0

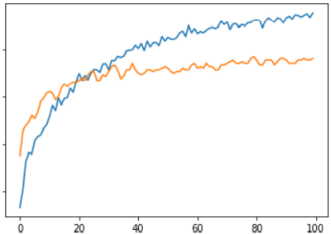
Konvolutivna

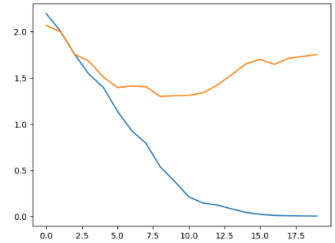
Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 256, 256, 32)	896
max_pooling2d (MaxPooling2D)	(None, 128, 128, 32)	0
conv2d_1 (Conv2D)	(None, 128, 128, 32)	9248
max_pooling2d_1 (MaxPooling2D)	(None, 64, 64, 32)	0
conv2d_2 (Conv2D)	(None, 64, 64, 32)	9248
max_pooling2d_2 (MaxPooling2D)	(None, 32, 32, 32)	0
dropout_3 (Dropout)	(None, 32, 32, 32)	0
flatten (Flatten)	(None, 32768)	0
dense_4 (Dense)	(None, 64)	2097216
dense_5 (Dense)	(None, 10)	650

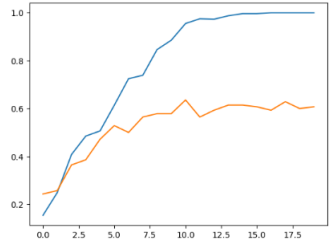
Total params: 2,117,258
Trainable params: 2,117,258
Non-trainable params: 0

Rezultati


2.0
1.5
1.0
0.5
0
20
40
60
80
100


0.8
0.6
0.4
0.2
0
20
40
60
80
100


2.0
1.5
1.0
0.5
0.0
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5


1.0
0.8
0.6
0.4
0.2
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5

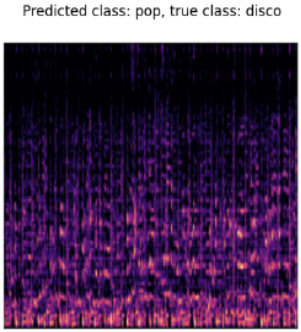
Matrice konfuzije

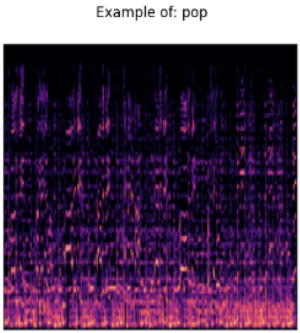
za element a_{ij} važi: stvarna klasa je i, predviđena klasa je j

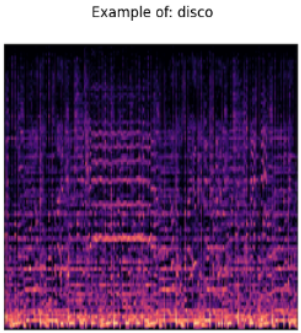
```
([[26,  0,  1,  1,  0,  0,  2,  0,  0,  0],
  [ 0, 29,  0,  0,  0,  1,  0,  0,  0,  0],
  [ 1,  0, 23,  0,  0,  0,  0,  1,  2,  3],
  [ 0,  0,  0, 17,  3,  0,  0,  5,  2,  3],
  [ 1,  1,  0,  0, 21,  0,  0,  0,  7,  0],
  [ 2,  1,  0,  0,  0, 27,  0,  0,  0,  0],
  [ 1,  0,  0,  3,  1,  0, 25,  0,  0,  0],
  [ 0,  0,  0,  1,  0,  0,  0, 26,  3,  0],
  [ 1,  1,  2,  0,  0,  1,  1,  0, 22,  2],
  [ 5,  0,  3,  5,  0,  2,  1,  0,  0, 14]])
```

```
([[20,  0,  2,  1,  1,  1,  1,  0,  3,  1],
  [ 0, 28,  0,  0,  0,  1,  0,  0,  0,  1],
  [ 2,  1, 16,  2,  0,  5,  0,  0,  1,  3],
  [ 0,  0,  2, 20,  3,  1,  1,  1,  0,  2],
  [ 0,  0,  0,  4, 20,  0,  0,  2,  3,  1],
  [ 5,  1,  3,  1,  0, 20,  0,  0,  0,  0],
  [ 0,  0,  1,  1,  0,  1, 26,  0,  0,  1],
  [ 1,  0,  1,  5,  2,  1,  0, 17,  0,  3],
  [ 1,  0,  1,  1,  5,  1,  0,  2, 18,  1],
  [ 5,  0,  1,  8,  0,  1,  2,  2,  1, 10]])
```

Pogrešno klasifikovane instance (1)







Pogrešno klasifikovane instance (2)

- biblioteka IPython
- display(Audio(file_path))

True class: jazz
Predicted class: classical

True class: pop
Predicted class: disco

Hvala na pažnji :)