

# Lumen – Data Science 2023

## Dokumentacija

# Sadržaj

Uvod.....	3
Cilj.....	3
Dataset.....	3
Pretprocesiranje podataka.....	3
Ekstrakcija značajki.....	3
Normalizacija podataka.....	3
Proširivanje podataka.....	4
Razvoj modela.....	4
Obuka modela:.....	5
Procjena modela.....	5
Zaključak.....	5

# Uvod

Audio klasifikacija je izazovan zadatak koji podrazumijeva identifikaciju i kategorizaciju audio signala na temelju njihovih karakteristika. U ovom projektu cilj nam je izgraditi model strojnog učenja koji može točno klasificirati audio signale u različite glazbene instrumentne kategorije na temelju njihovih spektrograma. Spektrogrami su vizualne reprezentacije frekvencijskog sadržaja audio signala tijekom vremena. Za obuku našeg modela koristit ćemo konvolucijske neuronske mreže (CNN), koje su se pokazale vrlo učinkovitima u zadacima klasifikacije zvuka.

## Cilj

Cilj ovog projekta je izgraditi model klasifikacije koji može točno klasificirati zvukove glazbenih instrumenata u različite kategorije na temelju njihovih spektrograma. Koristit ćemo različite algoritme strojnog učenja kako bismo postigli ovaj cilj, ali ćemo se prije svega usredotočiti na tehnike dubokog učenja korištenjem CNN arhitekture.

## Dataset

Koristit ćemo IRMAS skup podataka, koji je široko korišten skup podataka zvukova glazbenih instrumenata stvoren na Međunarodnoj konferenciji o pretraživanju glazbene informacije (ISMIR). Skup podataka sadrži više od 6700 audio datoteka 11 različitih glazbenih instrumenata. Skup podataka dostupan je za besplatno preuzimanje s sljedeće poveznice: <https://www.upf.edu/web/mtg/irmas>

## Pretprocesiranje podataka

Prije obuke modela potrebno je predobraditi podatke kako bi bili prikladni za naš algoritam strojnog učenja. Koraci predobrade uključuju:

### Ekstrakcija značajki

Koristit ćemo Mel spektrogram kao našu značajku, što je često korištena tehnika ekstrakcije značajki za audio signale. Mel spektrogrami izvedeni su iz Fourierove transformacije audio signala i pružaju vizualnu reprezentaciju frekvencijskog sadržaja audio signala tijekom vremena. Izvući ćemo 224 x 224 Mel spektrograma za svaku audio datoteku. Pokazano je da su Mel spektrogrami vrlo učinkoviti u zadacima klasifikacije zvuka, posebno kada se koriste s CNN-ovima.

### Normalizacija podataka

Normalizirat ćemo podatke skaliranjem između 0 i 1. To je bitan korak kako bi se osiguralo da naš model može učiti iz podataka na učinkovit način i donositi točne predikcije. Normalizacija pomaže u sprečavanju utjecaja razlika u veličini podataka na model.

## Proširivanje podataka

Proširivanje podataka je proces generiranja dodatnih podataka za obuku izvornih podataka primjenom transformacija poput dodavanja šuma, pomaka u visini ili promjene tempa. Proširivanje podataka pomaže u sprečavanju pretreniranja i poboljšava generalizaciju našeg modela. Primijenit ćemo tehnike proširivanja podataka poput nasumičnog pomaka u visini i ubrizgavanja bijelog šuma. Također ćemo nasumično obrezati Mel spektrograma kako bismo uveli varijabilnost u podatke za obuku.

## Razvoj modela

Razvit ćemo CNN model za klasifikaciju zvukova glazbenih instrumenata na temelju njihovih spektrograma. CNN-ovi su modeli dubokog učenja koji su posebno dizajnirani za rad s dvodimenzionalnim podacima poput slika i spektrograma. CNN model sastoji se od nekoliko konvolucijskih slojeva, nakon kojih slijede slojevi grupiranja i potpuno povezani slojevi. Konačni sloj modela će imati softmax aktivaciju kako bi se proizvela vjerojatnosna distribucija preko kategorija instrumenata.

Arhitektura CNN modela će se sastojati od sljedećih slojeva:

1. Konvolucijski sloj: Prvi sloj CNN-a će imati 32 filtra s veličinom jezgre  $3 \times 3$ . Koristit će se ReLU aktivacijska funkcija. Konvolucijski sloj će izvući značajke iz ulaznih spektrograma.
2. MaxPooling sloj: Izlaz iz konvolucijskog sloja bit će proslijeđen MaxPooling sloju s veličinom bazena  $2 \times 2$ . MaxPooling smanjuje prostorne dimenzije ulaza tako što uzima maksimalnu vrijednost u svakom bazenu.
3. Konvolucijski sloj: Dodat ćemo još jedan konvolucijski sloj s 64 filtera i veličinom jezgre  $3 \times 3$ . Ovaj će sloj također koristiti ReLU aktivaciju.
4. MaxPooling sloj: Izlaz iz drugog konvolucijskog sloja bit će proslijeđen još jednom MaxPooling sloju s veličinom bazena  $2 \times 2$ .
5. Sloj razrješavanja: Izlaz iz MaxPooling sloja bit će rastavljen kako bi se stvorio jednodimenzionalni vektor.
6. Potpuno povezani sloj: Dodat ćemo potpuno povezani sloj s 128 neurona i ReLU aktivacijom. Ovaj će sloj naučiti visoko razinske značajke iz ulaznih spektrograma.
7. Izlazni sloj: Konačni sloj CNN-a imat će 11 neurona s sigmoid aktivacijom, jedan za svaku kategoriju instrumenata. Sigmoid aktivacija proizvodi vjerojatnost za svaki instrument.

Kompilirat ćemo model korištenjem kategoričke križne entropije kao gubitka i Adam optimizatora. Također ćemo pratiti metriku točnosti tijekom obuke kako bismo evaluirali performanse modela.

```
model = Sequential()
model.add(Conv2D(32, (3, 3), activation='relu', input_shape=(224,
224, 3)))
model.add(MaxPooling2D(pool_size=(2, 2)))
model.add(Conv2D(64, (3, 3), activation='relu'))
model.add(MaxPooling2D(pool_size=(2, 2)))
model.add(Flatten())
model.add(Dense(128, activation='relu'))
model.add(Dense(11, activation='sigmoid'))

model.compile(loss='binary_crossentropy', optimizer='adam',
metrics=['accuracy'])
```

## Obuka modela:

Obučit ćemo CNN model korištenjem preprocesiranog IRMAS skupa podataka za obuku i IRMAS\_Validation\_Data za validaciju. Model ćemo trenirati 20 epoha s veličinom grupe (engl. batch size) od 32. Tijekom obuke koristit ćemo ranu zaustavu (engl. early stopping) kako bismo spriječili prenaučенost. Također ćemo spremiti najbolje težine modela na temelju točnosti validacije.

## Procjena modela

Procijenit ćemo performanse naučenog modela na testnom skupu podataka, koji nismo koristili tijekom treninga ili validacije. Izračunat ćemo mjere kao što su točnost, preciznost, odziv i F1-score kako bismo procijenili performanse modela. Također ćemo vizualizirati matricu zabune kako bismo vidjeli koje instrumente model ima poteškoća u razlikovanju.

## Zaključak

U ovom projektu razvili smo CNN model za klasifikaciju zvukova glazbenih instrumenata na temelju njihovih spektrograma. Za trening i validaciju koristili smo IRMAS skup podataka te postigli visoku točnost u klasifikaciji različitih kategorija instrumenata. Ovaj model mogao bi se koristiti u raznim aplikacijama, poput sustava za preporuku glazbe, automatskih transkripcijskih sustava i softvera za proizvodnju glazbe.