Klasifikacija glazbe po žanru

Šime Batović Mislav Vučković
Andrija Mandić Marko Jukić

Prirodoslovno-matematički fakultet — Matematički odsjek

24. lipnja 2020.

Sadržaj

Motivacija i ciljevi

Analiza i klasifikacija glazbe danas su dobro istraživana područja. Popularne aplikacije poput *Spotify* i *Google Play Music* se već dugi niz godina bave ovim problemom.

Cilj istraživanja: na problemu klasifikacije žanrova usporediti **točnosti** klasičnih metoda strojnog učenja i metode dubokog učenja koristeći konvolucijske neuronske mreže.

Skup podataka

Koristili smo FMA dataset (`fma_small`):

- 8000 isječaka pjesama duljine 30 sekundi,
- mp3 format,
- 1000 pjesama za svaki od 8 žanrova.

Žanrovi: Experimental, Hip-Hop, Rock, Pop, Folk, Electronic, Instrumental i International.

Pristup rješavanju problema

Klasifikaciji glazbe po žanru:

1. izračunati razne spektralne i ritamske značajke za svaku pjesmu, te pomoću njih i klasičnih metoda stojnog učenja pokušati odrediti žanr,
2. reprezentirati pjesme mel-spektogramima i iskoristiti konvolucijsku neuronsku mrežu za klasifikaciju.

Korištena tehnologija

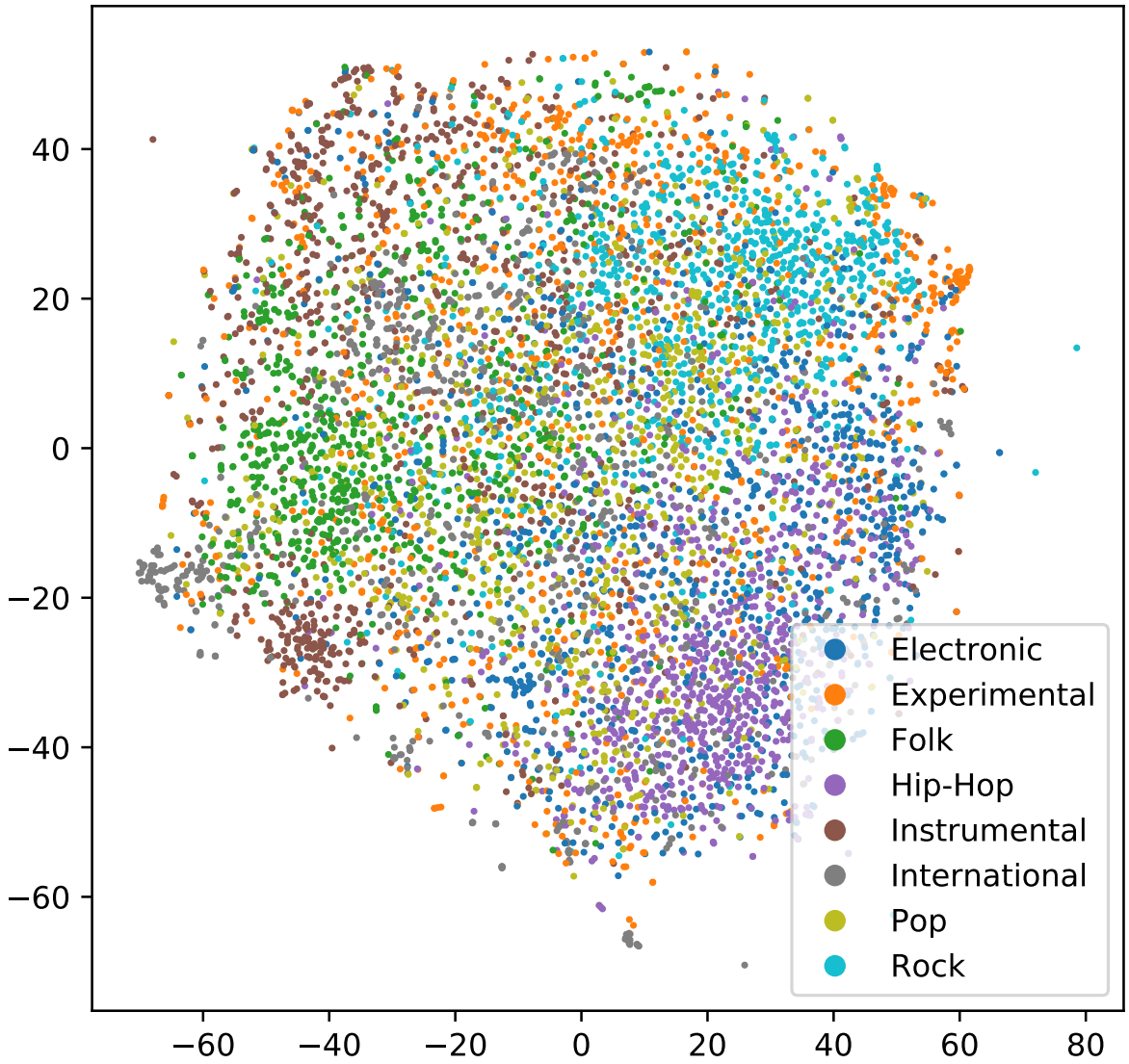
- Python
- Treniranje modela na *Kaggle*-u, koristeći grafičke procesore za ubrzanje
- Tesla P100 grafički procesor sa 16GB memorije

Klasične metode

Koristili smo dvije vrste značajki za klasifikaciju: **spektralne** (*special centroid*, *special bandwidth*, *special contrast*...) i **ritamske** (tempo). Ukupno smo izračunali 380 značajki za svaku pjesmu.

Za računanje svih značajki koristili smo Python paket *LibROSA*.

Vizualizacija pomoću t-SNE



Klasične metode

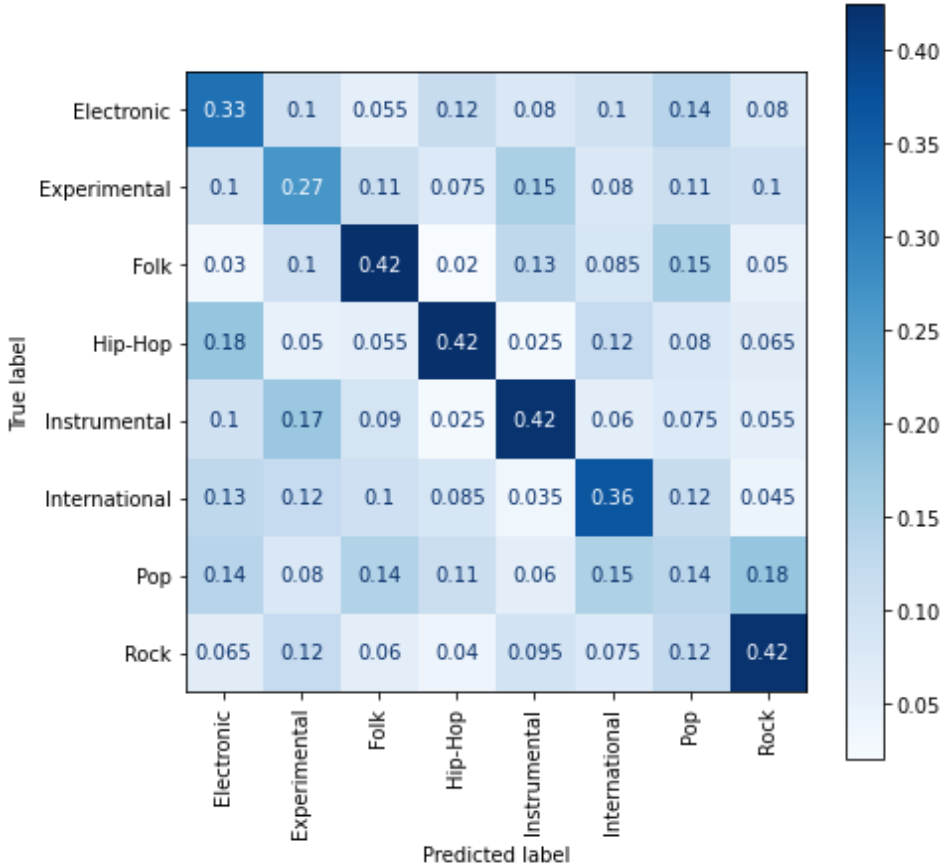
Istražili smo šest klasičnih modela strojnog učenja:

- Stablo odlučivanja
- Slučajna šuma
- Logistička regresija
- Metoda potpornih vektora (SVM)
- AdaBoost
- XGBoost

Skup značajki podijelili smo na train (80%) i test (20%). Za svaki model na train skupu smo isprobali veliki broj parametara pomoću `GridSearchCV` i za model s najboljim parametrima odredili točnost na testnim podacima.

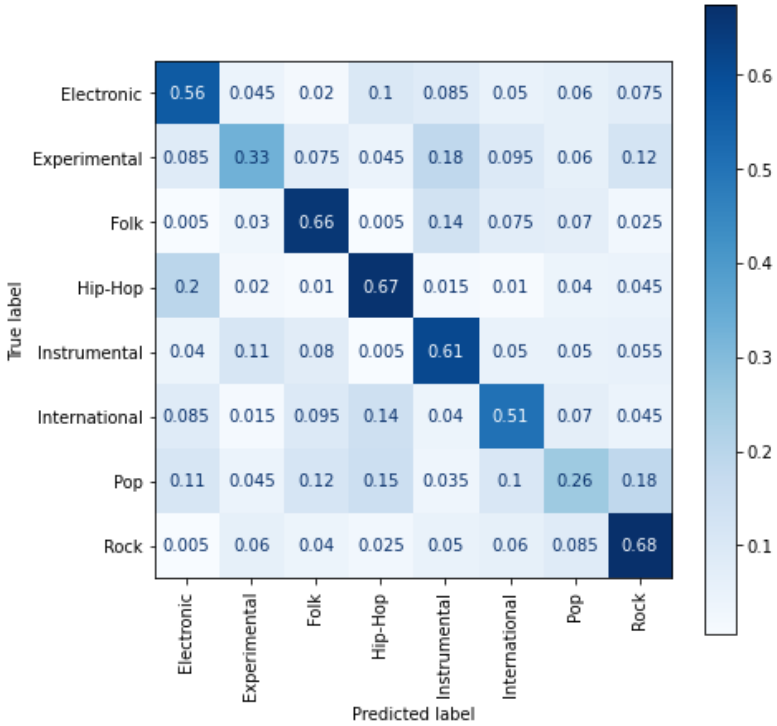
Stablo odlučivanja

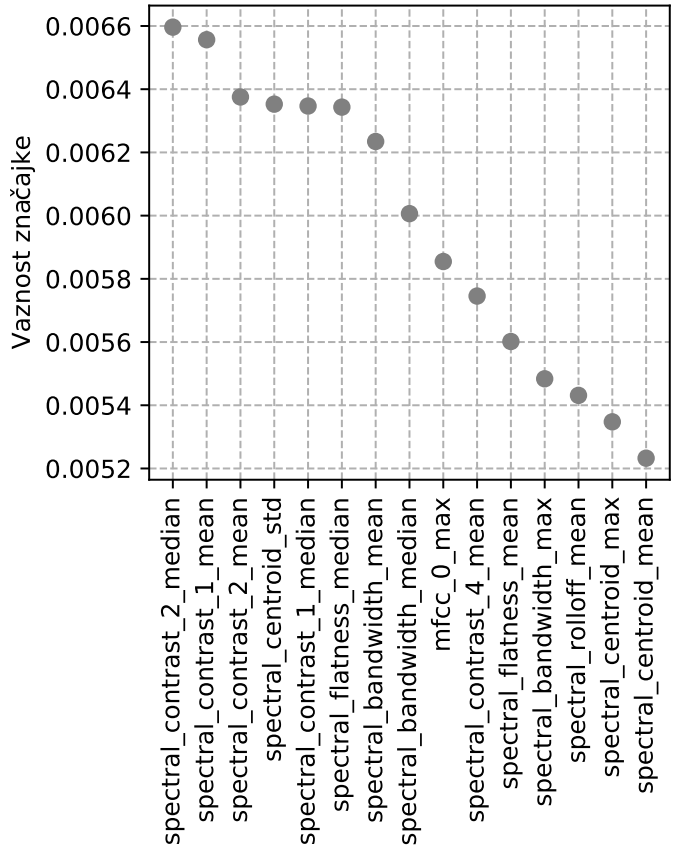
Stabla odlučivanja su se pokazala kao dosta nestabilan model. Pretraživanjem velikog broja parametara pomoću `GridSearchCV` postigli smo maksimalnu točnost od samo 34.75% na testnim podacima.



Slučajna šuma

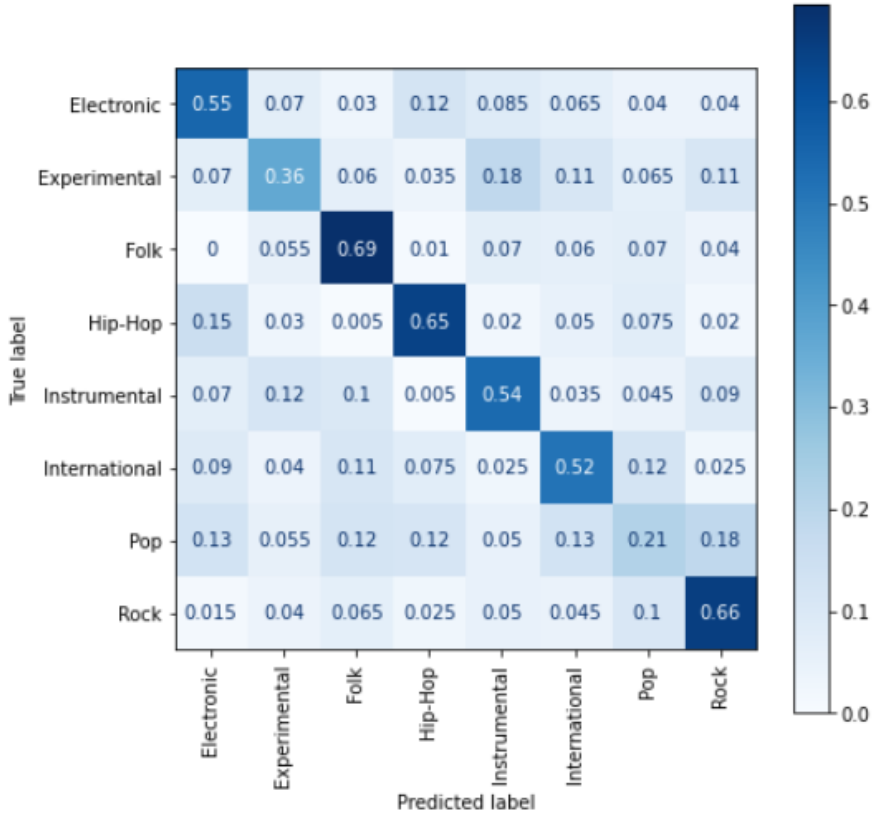
Uz neograničenu maksimalnu dubinu stabala, funkciju razdvajanja na temelju gini indexa, te 1000 stabala u šumi, uspili smo postiči točnost od 53.25% na testnim podacima.





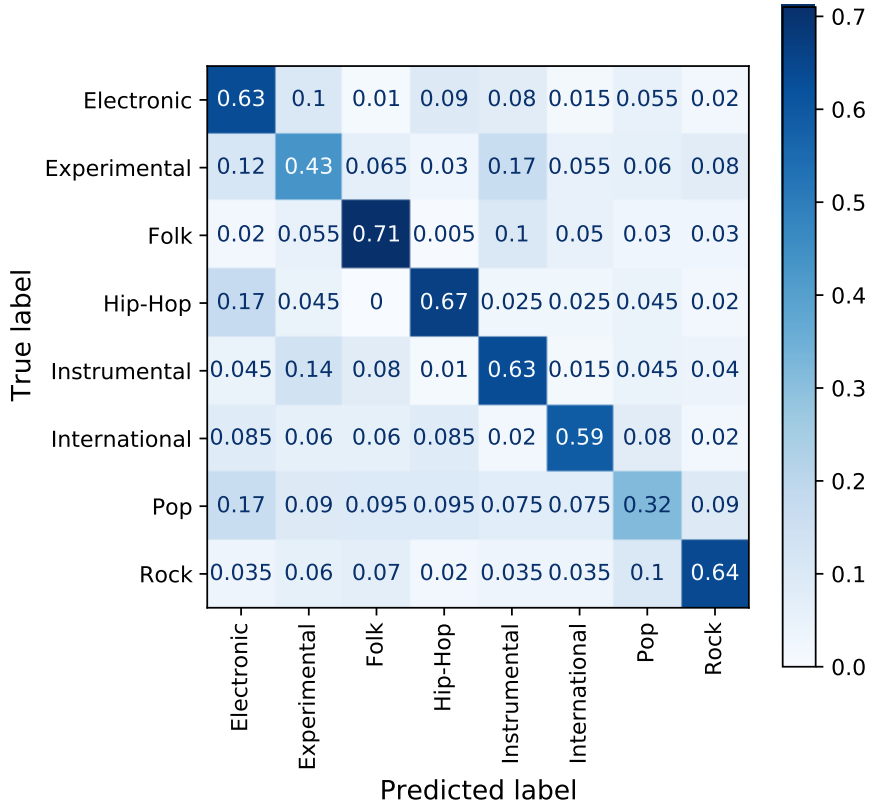
Logistička regresija

Pomoću `GridSearchCV` ponovo tražimo najbolji model. Za odabir značajki nam služi prethodno izračunat model slučajne šume. Uz maksimalan broj iteracija postavljen na 1000, model postiže točnost od 52.25% na testnim primjerima.



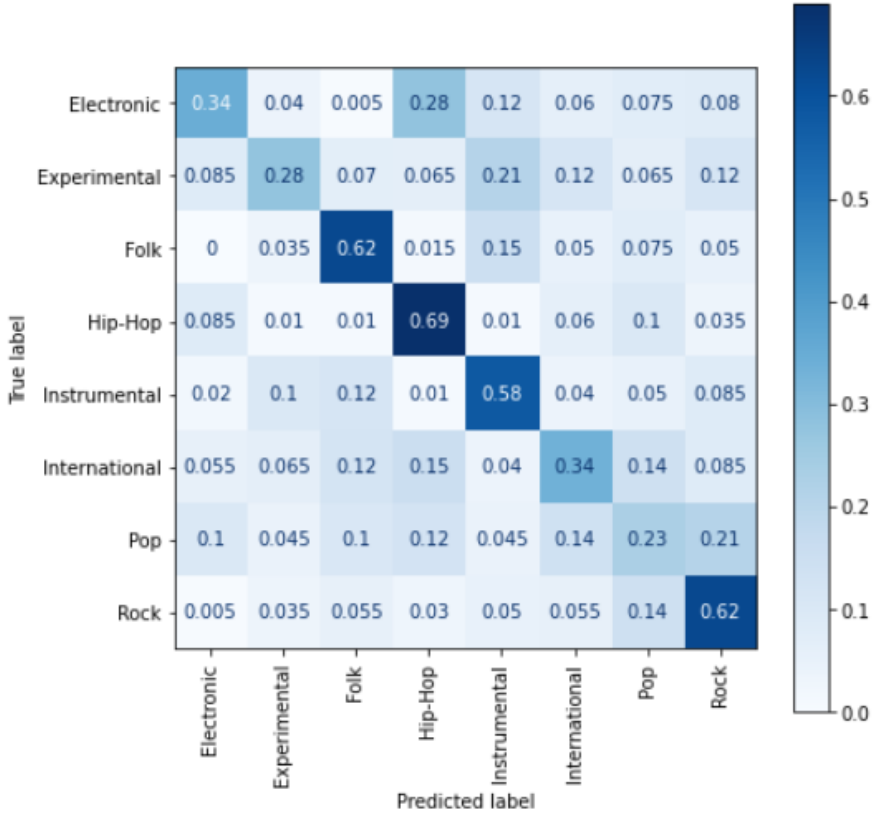
Metoda potpornih vektora - SVM

Ponovo za odabir značajki koristimo model slučajne šume. RBF kernel služi za mapiranje u veće dimenzije te žrtvujemo veličinu margine za što bolju klasifikacijsku točnost. Model postiže točnost od 57.63%.



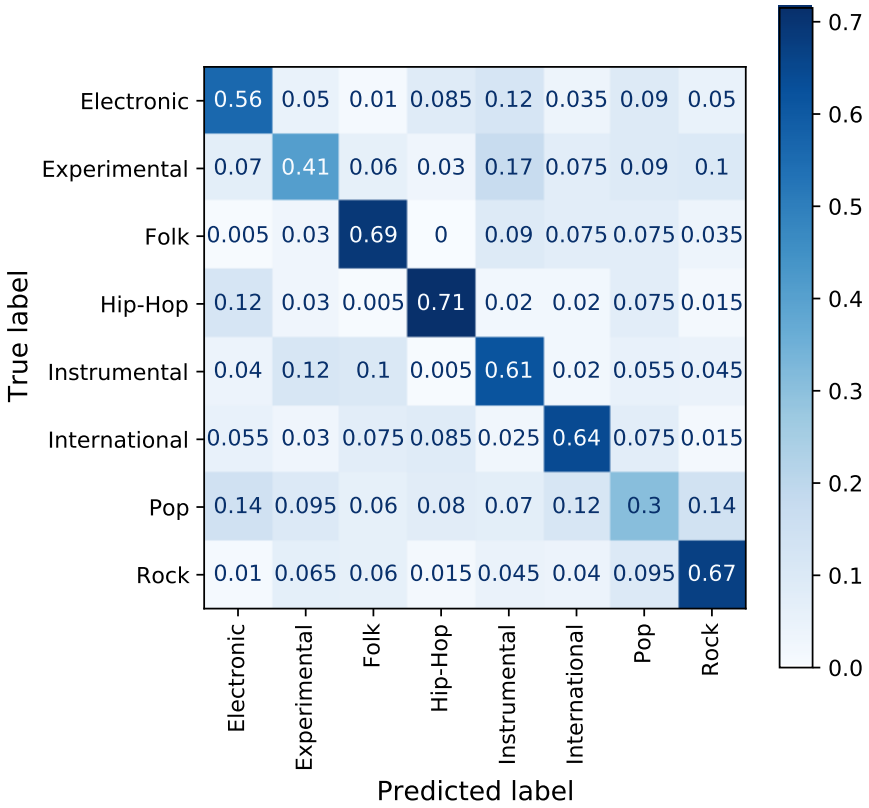
AdaBoost

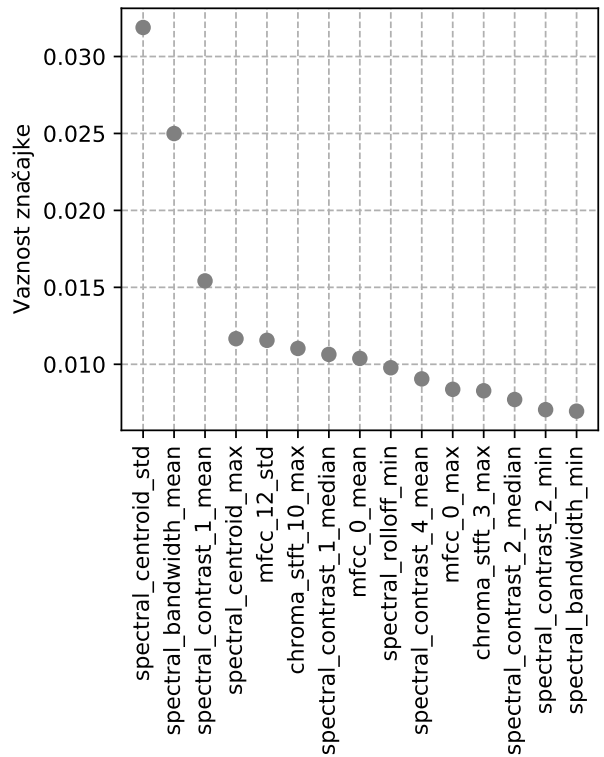
Koristimo AdaBoost algoritam s povećanim brojem estimatora i smanjenim `learning_rate` koji određuje koliko svaki model pridonosi postojećem. Postignuta točnost na testnim primjerima je loša, samo 46.31%.



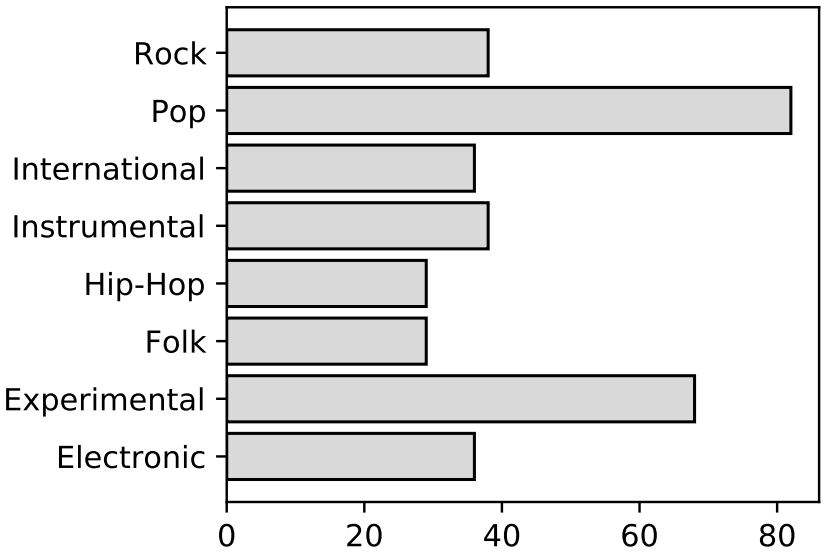
XGBoost

XGBoost se zbog svoje robustnosti na outliere pokazao kao najbolja metoda. Postavljanjem broja stabala na 180 i `learning_rate` na 0.25 dalo nam je najbolju točnost od 57.50% na testnim primjerima.





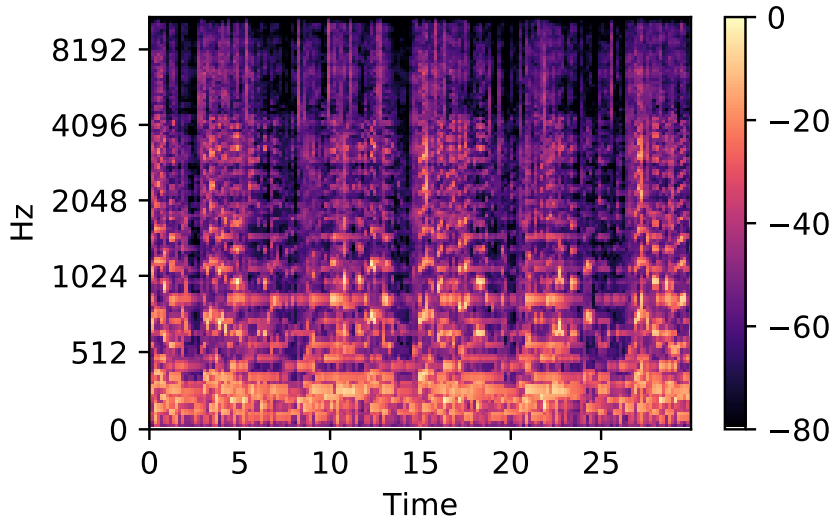
Analiza modela



Klasičnim metodama dobili smo uglavnom očekivane rezultate kad usporedimo s prethodnim istraživanjima na istom datasetu. Na slici lijevo vidimo koliki je broj pjesama po žanru koje niti jedan model nije točno klasificirao.

Konvolucijske neuronske mreže

Za opisivanje svake pjesme koristili smo mel-spektogram, dvodimenzionalni graf koji prikazuje jačinu frekvencija u ovisnosti o vremenu.



Slika: Mel-spektogram pjesme žanra Folk

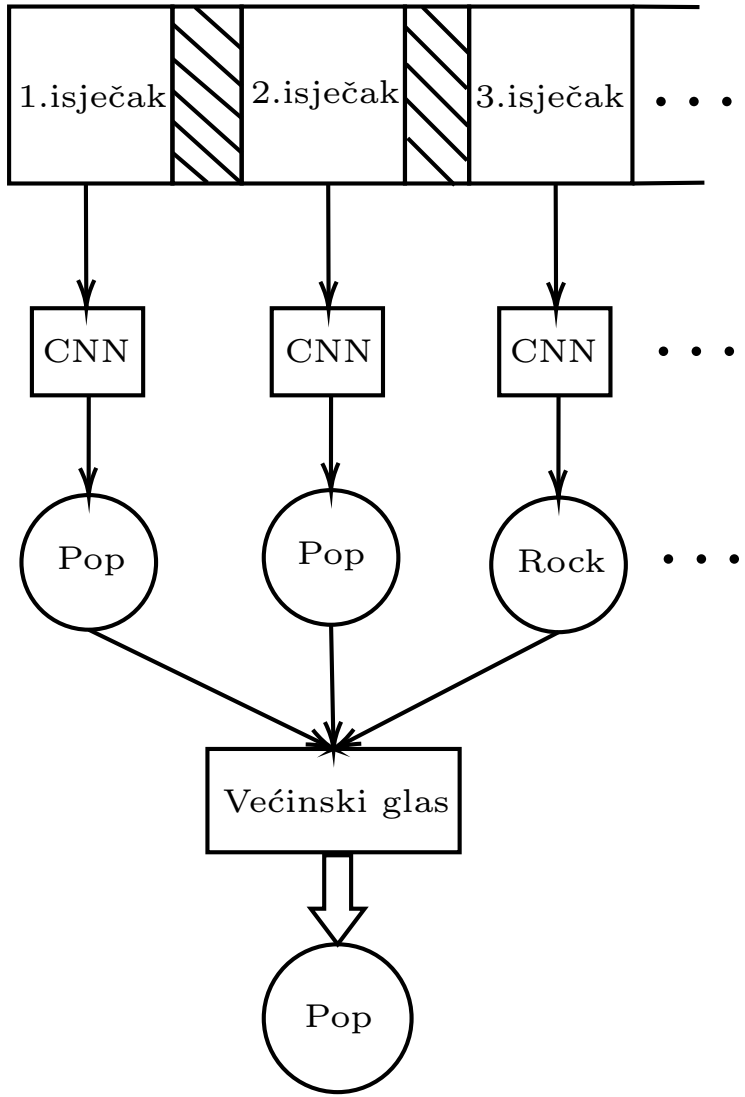
Metoda

Originalni mel-spektogrami većinom dimenzija 128×1291.

Podjela na manje slike dimenzija 128×16 na kojima smo učili neuronsku mrežu.

Konačni model je *metoda većinskog glasa*.

Metoda većinskog glasa



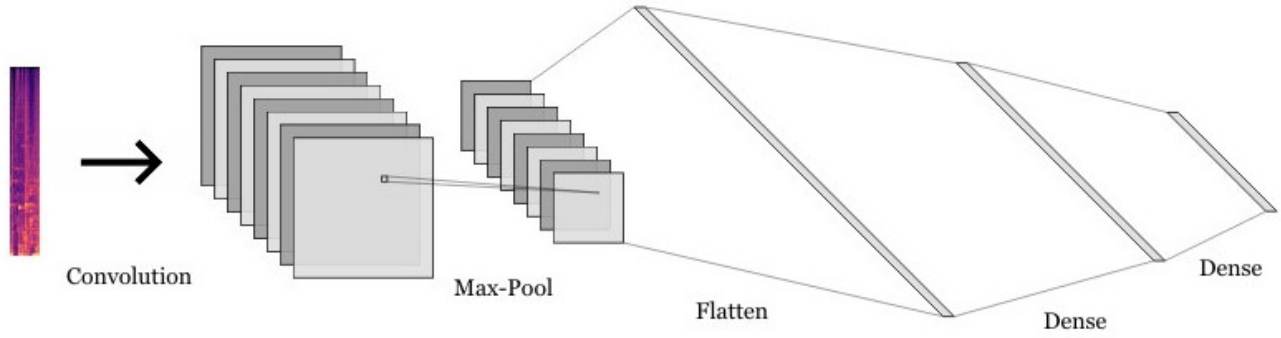
Metoda

Podjela podataka: *train* 72%, *validation* 8% i test 20%.

Prilikom učenja koristili smo *Stohastic gradient descent* s parametrima $learning\_rate = 0.001$ i $momentum = 0.9$ te je tokom učenja korišten $batch\_size = 256$.

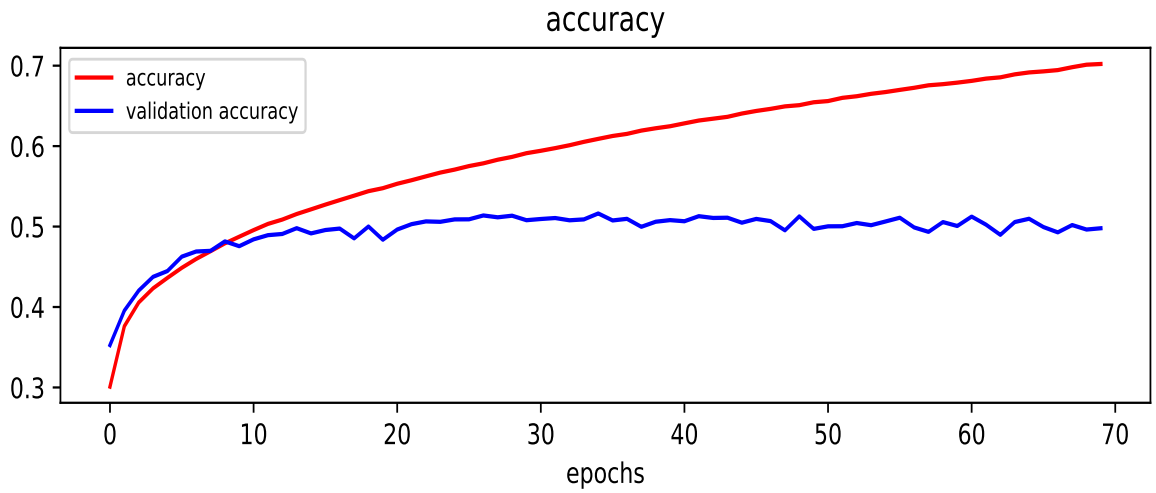
Korištene razne arhitehture (njih 11). Osnovna ideja: nizanje konvolucijskih slojeva zajedno s *Max-Pooling* slojem, nakon čega slijedi niz potpuno povezanih slojeva.

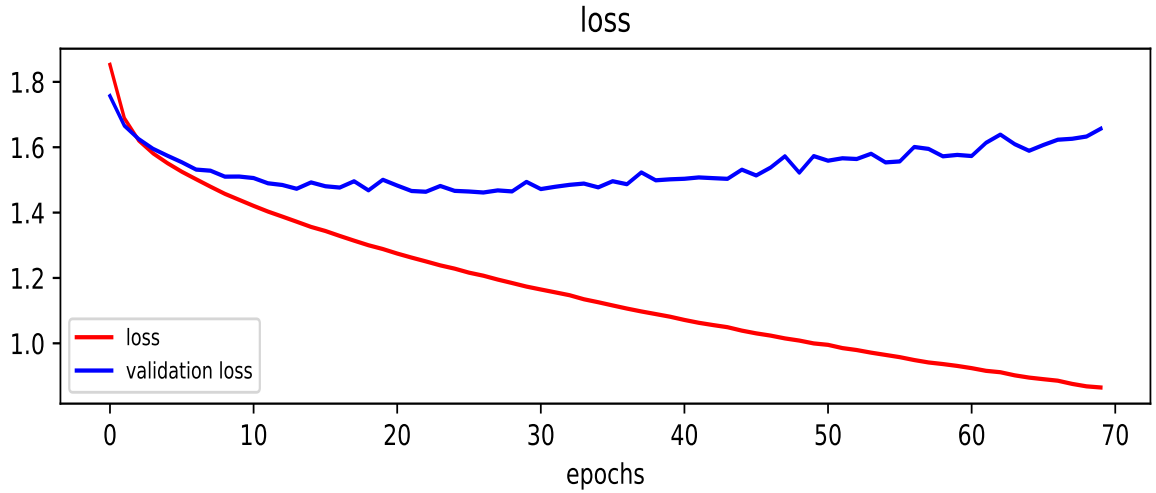
Arhitektura neuronske mreže




```
Sequential([
        InputLayer(input_shape=(height, width, 1)),
        Conv2D(128, (5, 5), activation='relu',
    padding='same', strides=1),
        MaxPooling2D(pool_size=(2, 2), strides=2),
        Flatten(),
        Dense(32, activation="relu"),
        Dense(8, activation="softmax")
    ]),
```

Krivulja učenja





Prikaz rezultata

Usporedba točnosti modela mjerenih na testnim primjerima:

Model	Točnost
Stablo odlučivanja	34.75%
Slučajna šuma	53.25%
Logistička regresija	52.25%
Metoda potpornih vektora	57.63%
AdaBoost	46.31%
XGBoost	57.50%
CNN	58.69%

Osvrt na druge pristupe

Tim s pekinškog sveučilišta, na istom *datasetu*, koristeći CNN postigao točnost 59.4% (*metoda većinskog glasa*, *data augmentation*, *Conv1D layeri*).

Korištenjem rezidualne neuronske mreže – *ResNet*-a, SVM-om kao *stacking classifier*-om umjesto metode većinskog glasa, postigli su točnost od 66.3%.

Nastavak istraživanja

- Istraživanje novih arhitektura konvolucijske neuronske mreže.
- Učenje na većem skupu podataka (hardverska ograničenja).
- Korigiranje *overlap*-a kod rezanja slika.
- Korištenje *data augmentation* radi povećanja volumena dataseta i smanjivanja *overfitting*-a
- Korištenje drugih podataka o pjesmama, poput podžanrova.

Hvala na pažnji!