

# Analiza audio scene

Nikola Pupovac, IN31/2018, [nikola\\_s99@live.com](mailto:nikola_s99@live.com)

Marko Vujić, EE 229/2018, [vujiemarko222@gmail.com](mailto:vujiemarko222@gmail.com)

## I. UVOD

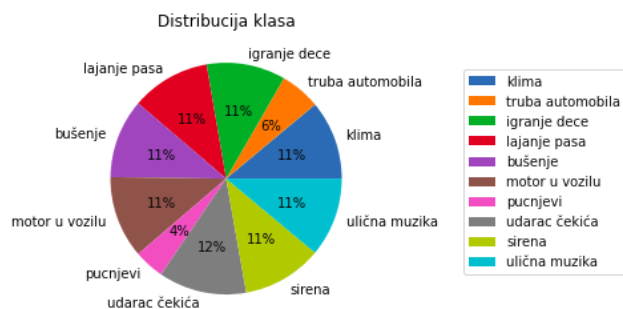
Audio-vizuelna analiza scene u cilju detekcije anomalija (sudara, pucnjeva i sl.) kao i u cilju ostvarenja pametnih gradova, danas je jedna od izuzetno aktuelnih oblasti istraživanja. Cilj ovog istraživanja jeste napraviti model koji što tačnije vrši klasifikaciju zvuka, tj. predviđa da li je zvuk bio zvuk sudara, plača deteta i sl. na osnovu akustičkih obeležja tog zvuka.

## II. BAZA PODATAKA

Baza podataka predstavlja kolekciju akustičnih obeležja zvukova kao i obeležja kom tipu zvuka ona pripadaju (zvukova iz urbane sredine, poput saobraćaja, sirene, ptica, govora ljudi i dr.).

Zvuk	Broj uzoraka	procenatualno
Klima	600	11%
Truba Automobila	306	6%
Igranje dece	600	11%
lajanje pasa	600	11%
Bušenje	600	11%
Motor u vozilu	624	11%
Pucnjevi	230	4%
Udarac čekića	668	12%
Sirena	607	11%
Ulična muzika	600	11%

Baza podataka je data u vidu csv fajla i broji 5435 uzoraka od kojih se može zapaziti da baza sadrži najviše uzoraka udarca čekićem, a najmanje zvukova pucnjeva, kao i trube automobila.



Slika 1. Distribucija zvukova u bazi podataka.

Baza sadrži tačno 167 obeležja 166 obeležja realnog numeričkog tipa koje predstavljaju akustična obeležja zvuka, kao i jedno kategoričko obeležje koje nam govori o tipu zvuka kome uzorak priada.

## III. ANALIZA PODATAKA

Baza podataka nema nedostajućih vrednosti, te nije potrebno nadopuniti ili odbaciti podatke. Pošto baza predstavlja akustična obeležja zvukova, sva obeležja su zadržana.

Vrednosti za različita obeležja znatno variraju po vrednosti, gde se vrednosti nekih obeležja kreću od 0 do 1, dok se je kod jednog obeležja najveća vrednost 2732. Imajući to u vidu, neophodno izvršiti normalizaciju obeležja, kako mnogi algoritmi ne bi dali prioritet određenim obeležjima zbog njihove vrednosti.

Klase zvukova su podeljene prema taksonomiji zvukova u urbanim sredinama, za više o tome pogledati sliku 3 i pročitati izvor naveden pod [2].

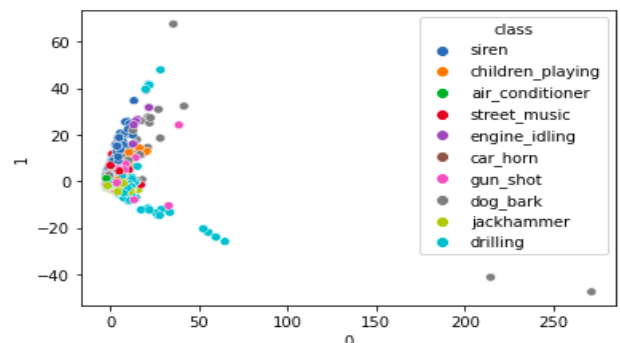
## IV. PRIPREMA PODATAKA

Podaci su za početak podeljeni u skupove za treniranje i testiranje modela, gde skup za treniranje obuhvata 90% podataka, dok test set obuhvata ostatak podataka. Uzorci su odabrani nasumično, a vođeno je i računa o tome da distribucija podataka po klasama bude ravnomerna tako da i trening i test skup sadrže procentualno približno jednako uzoraka svakog tipa zvuka.

Kako je navedeno u analizi podataka, vrednosti po obeležjima se dosta razlikuju, te ih je neophodno normalizovati. Za normalizaciju korišćena formula:

$$Z = \frac{x - \mu}{\sigma}$$

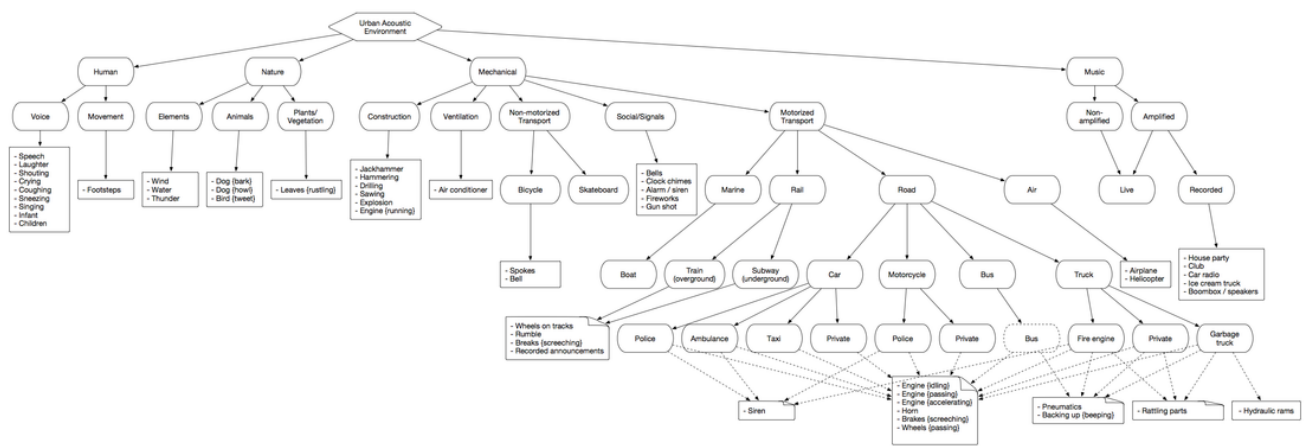
Zbog velike količine obeležja, izvršen je "PCA" na kopiji skupa za treniranje kako bi se broj obeležja smanjio na 2, te mogao vizuelizovati.



Slika 2. Distribucija uzoraka po PCA1 i PCA2.

[1] A Dataset and Taxonomy for Urban Sound Research by Justin Salamon, Christopher Jacoby and Juan Pablo Bello [http://www.justinsalamon.com/uploads/4/3/9/4/4394963/salamon\\_urbansound\\_acmmm14.pdf](http://www.justinsalamon.com/uploads/4/3/9/4/4394963/salamon_urbansound_acmmm14.pdf)

[2] S. R. Payne, W. J. Davies, and M. D. Adams. Research into the practical and policy applications of soundscape concepts and techniques in urban areas. DEFRA, HMSO, London, UK, 2009.



Slika 3. Taksonomija obeležja

Kao što je vidljivo iz slike 2, neophodno je ukloniti outlajere iz trening seta kako bi model bio uspešno obučen. Nakon uklanjanja outlajera, zadržano je 99.6% podataka, dok je najviše uklonjenih uzoraka pripadalo zvuku bušenja.

## V. ODABIR MODELA

Kako je traženo u zadatku, odabrana su 3 modela obučena na skupu za treniranje te je zatim svaki od modela ocenjen na test skupu. Kao ocena se koristio F1 mikro skor iz razloga što klase nisu ravnomerno raspoređene, te ocena tačnosti ne bi davala jasnu sliku uspešnosti modela, dok F1 skor kao rezultat harmonijske regresije osteljivosti i preciznosti daje jasnu sliku. Potom je izvršena redukcija dimenzionalnosti korišćenjem pa su modeli ponovo trenirani.

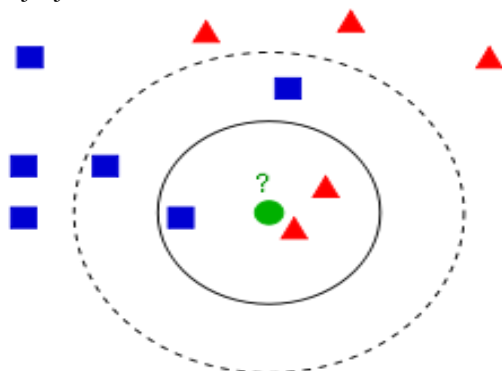
Za modele su odabrane KNN, Neuralne mreže, kao i Logistička regresija, a za redukciju dimenzionalnost odabrana je "PCA" sa brojem principijalnih komponenti takvim da sadrži 90% podataka.

### 1. KNN

Kako je traženo u zadatku, odabran prvi odabran model je KNN - k nearest neighbors (k najbližih suseda na srp.).

Ovaj algoritam spada u kategoriju algoritama nadgledanog učenja, što znači da koristi već poznate izlazne vrednosti.

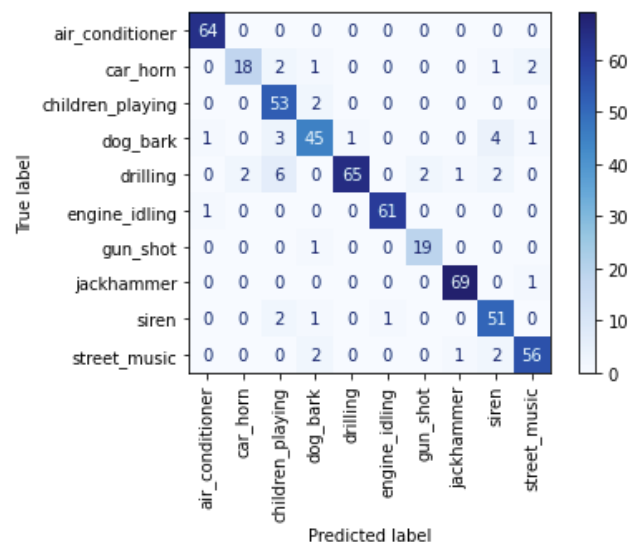
KNN vrši klasifikaciju uzorka na osnovu njemu "k" najbližih uzoraka. "k" predstavlja broj najbližih suseda od uzorka koji su predefinisani po takođe predefinisanim rastojanjem.



## Slika 4. Knn klasifikacija

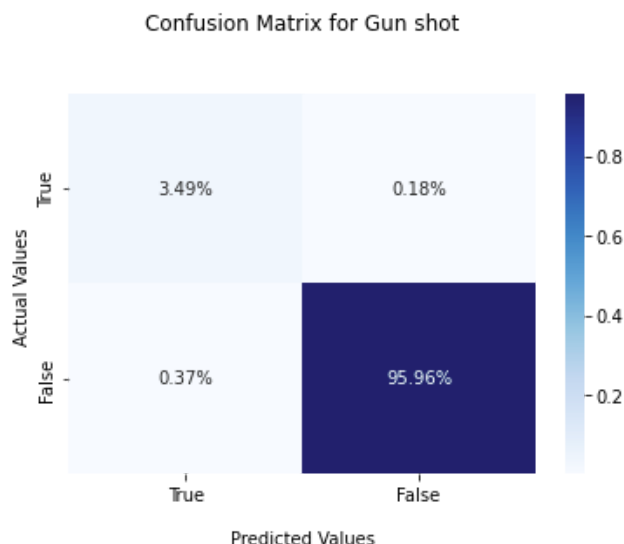
Algoritam je testiran sa više parametara "k", kao i "euclidean"(euklidsko na srp.) i "manhattan"(menhetn) merila rastojanja od suseda. Vršena je "Kfold" podela baze podataka, u kojoj su podaci pomešani, kojom je baza podeljena na 10 delova sa podjednakim brojem uzoraka svakog mesta porekla recepta. Na ovaj način se 10 puta uzima drugi deo baze podataka kao dev set, a ostatak korišćen za treniranje modela.

Najveća ocena tačnosti se pokazala korišćenjem parametra k=1 kao i euklidskog rastojanja koja je iznosila približno 0.922, dok je primenom redukcije dimezionalnosti ocena smanjena na 0.920.



Slika 5. Matrica konfuzije KNN

Matrica konfuzije svih klasa ne daje jasnu sliku uspešnosti modela, te je odlučeno da pored nje bude prikazana matrica konfuzije zvuka pucnja, kao najbitnijeg zvuka.

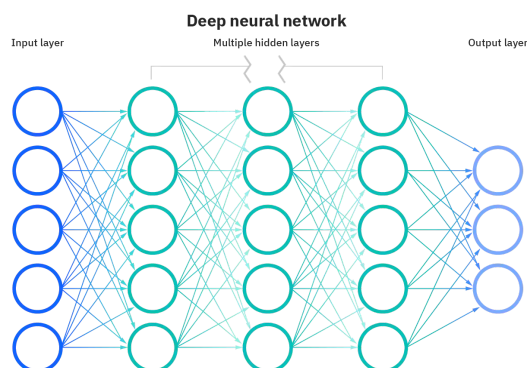


Slika 6. Matrica konfuzije za pucanj KNN modelom

## 2. Odabir modela 2

Kao drugi model, odabran je NN - Neural network

Ovaj algoritam takođe spada u kategoriju algoritama nadgledanog učenja. On se obučava tako što traži optimalne težine i između čvorova, kao i “bias” koje služe za računanje i konačno klasifikovanje.



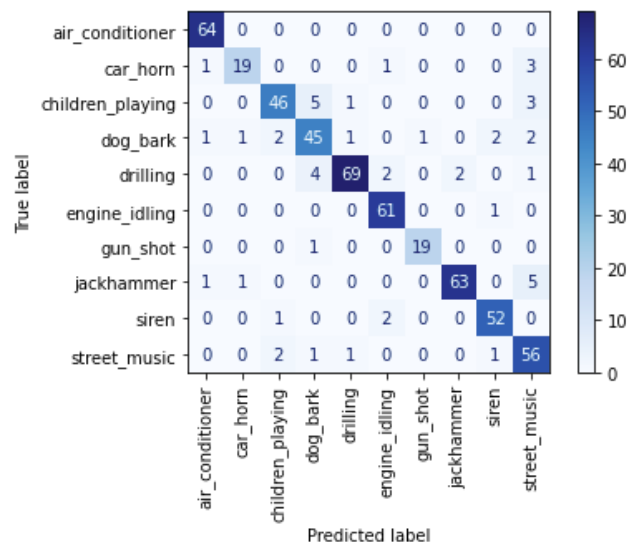
Slika 7. NN

Algoritam je testiran sa više različitih parametara, gde su se kao optimalni parametri pokazali:

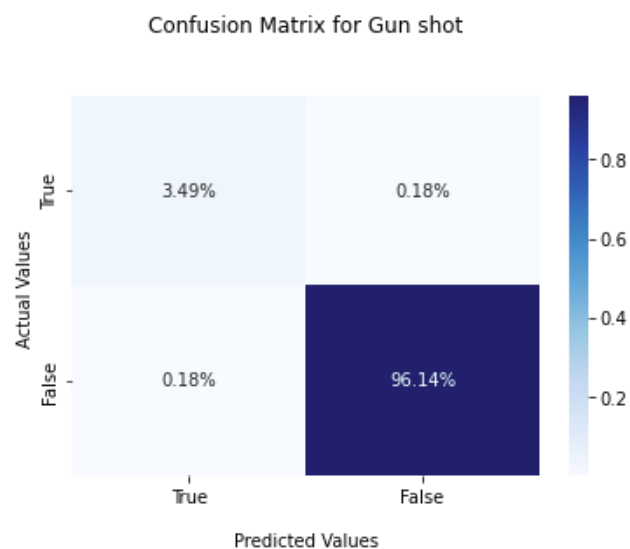
- Parametar učenja: adaptivan
- Slojevi: (128,64)

Algoritam ima manju ocenu od KNN algoritma, koja iznosi 0.908, dok se merilo performansi još znatno samnjilo na 0.878 nakon redukcije dimenzionalnosti.

Međutim ono što možemo zapaziti jeste da ovaj model bolje predviđa da bolje predviđa zvuk pucnja



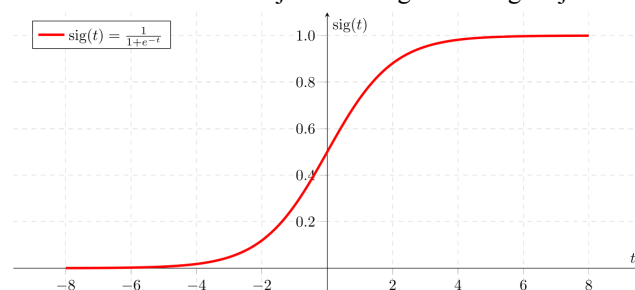
Slika 8. Matrica konfuzije NN modelom



Slika 9. Matrica konfuzije za pucanj NN modelom

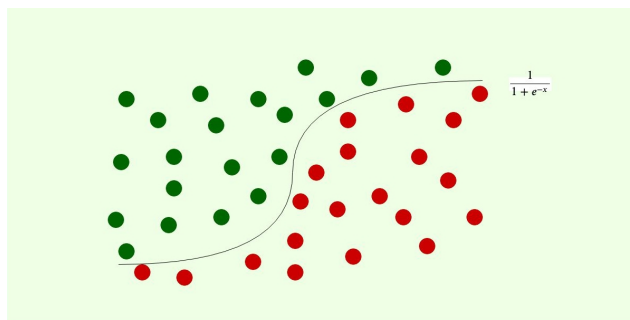
## 3. Odabir modela 3

Kao treći model odabran je model logističke regresije.



Slika 10. Logistička regresija

Model logističke regresije se obučava tako što traži optimalne parametre za fitovanje krivih koje mu služe za klasifikaciju.



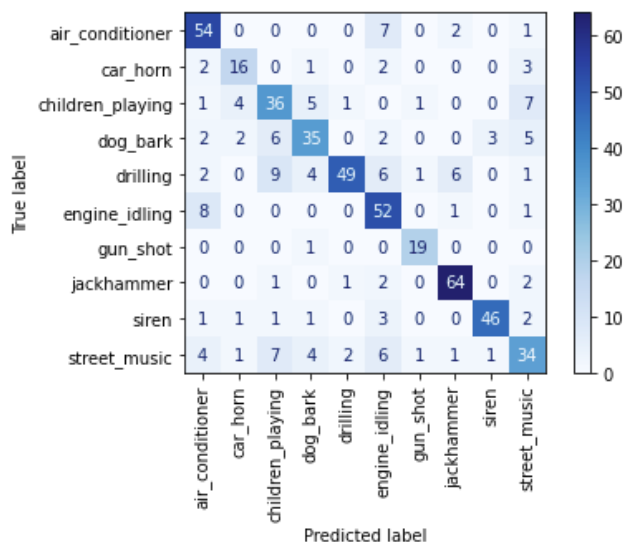
Slika 10. Logistička regresija - klasifikacija

## VI. POREĐENJE MODELA I ZAKLJUČAK

KNN se pokazao kao najbolji model za predikciju klasa, potom NN pa Logistička regresija. Osim bolje ocene, KNN algoritam se za ovako mali broj uzoraka pokazao kao najbrži 20 sekundi da se istrenira, za razliku od neuralnih mreža i logističke regresije kojima je trebalo znatno više, što ga u ovom slučaju čini najboljom opcijom.

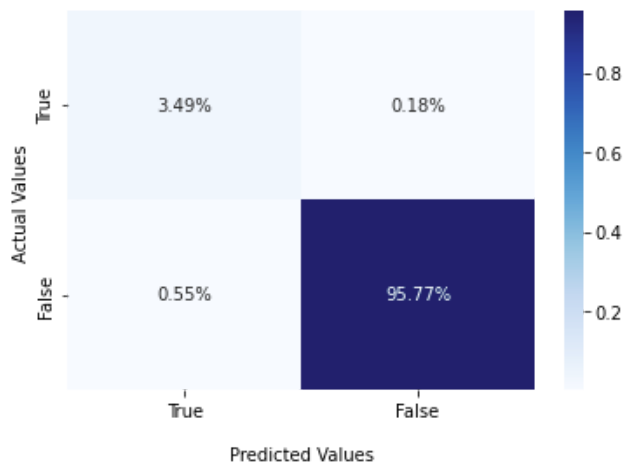
Algoritam je testiran sa više različitih parametara, gde su se kao optimalnan parametar za obučavanje pokazao “solver” “newton-cg”.

Algoritam ima znatno manju ocenu od KNN i NN algoritma, koja iznosi 0.744, dok se merilo performansi još znatno samnjilo na 0.687 nakon redukcije dimenzionalnosti.



Slika 11. Matrica konfuzije modelom logističke regresije.

Confusion Matrix for Gun shot



Slika 12. Matrica konfuzije za pucanj NN modelom