

Audio Classification

LUMEN Data Science

2023



Contents

1	O nama	3
2	Uvod u problem	3
2.1	Poslovni problem	3
2.2	Edukativni segment ovog izazova	4
3	Zadatak	5
3.1	Opis dataset-a	5
4	Format rješenja	7
5	Elementi ocjenjivanja	8

1 O nama

- Ime: Atomic Intelligence d.o.o.
- Adresa: Bencekovićeve 33, 10000 Zagreb
- Kontak osoba: Tomislav Križan
- e-mail: tkrizan@atmc.ai
- mobitel: +385 98 380 494

Atomic Intelligence d.o.o. je tvrtka sa sjedištem u Zagrebu (HR). primarno usmjerena na EU i SAD tržišta specijalizirana u područjima upravljanja podataka, integracije podataka, skladištenja podataka i poslovne inteligencije, implementacije podatkovnih aplikacija, rudarenje podataka, strojnog učenja i umjetne inteligencije u tradicionalnim analitičkim sustavima, ali i u sustavima koji koriste Big Data okruženja i platforme. Osim usluga implementacije rješenja unutar prije navedenih tematskih područja, Atomic Intelligence d.o.o. (skraćeno AI) nudi usluge iz područja strateškog IT savjetovanja, razvoja rješenja po narudžbi prilagođenog specifičnim potrebama korisnika i stručnosti u područjima poduzeća i rješenja arhitektura. AI zaposlenici imaju veliko i višegodišnje/desetljetno praktično iskustvo u različitim industrijama što može biti potkrijepljeno relevantnim certifikatima.

2 Uvod u problem

Glazba je prisutna u našim svakodnevnim životima, pojavljuje se u svim kulturama kroz povijest i neprestano se razvija. Shodno tomu razvijaju se i usluge vezane za glazbu. Veliki streaming servisi trude se što više poboljšati njihov korisnički doživljaj uređivanjem sadržaja i personalizacijom preporuka svojim korisnicima. Kako bi sve to učinili posjeduju moćne alate za glazbenu analizu i dohvaćanje raznih informacija o pjesmama. Jedna od najbitnijih takvih informacija je pojavljivanje raznih instrumenata. Mogućnost automatske klasifikacije instrumenta (također vokala) unutar pjesme omogućuje razne primjene poput lakšeg pretraživanja, točnijih preporuka te daljnje još detaljnije i kompleksnije analize. Iako ljudskom uhu detekcija instrumenata zvuči kao relativno lagan zadatak, automatizirati takav proces nije nimalo trivijalno. Visoka entropija informacija sadržanih u audio signalima, široki raspon izvora, postupak miksiranja i teškoća analitičkog opisa zahtjeva nešto kompleksnija rješenja.

2.1 Poslovni problem

Prepoznavanje glazbenih instrumenata ima različite primjene poput:

- Automatsko anotiranje glazbenih multimedijских podataka

- Transkripcija glazbenih izvedbi u svrhu podučavanja
- Sustav koji može razumijeti glazbu dovoljno kako bi u stvarnom vremenu surađivao s korisnikom u izvedbi
- etc.

Klasifikacija instrumenata i ljudskog glasa samo po sebi već je koristan dio informacije, međutim ona se koristi i u drugim problemima poput separacije određenih dijelova pjesme iz miksanih audio datoteka, generiranje glazbenih točaka, klasifikacija žanra određene pjesme i općenito u mnogim klasifikacijama audio signala. Isto tako detekcija i prepoznavanje različitih zvukova vrlo su važni u mnogim područjima koja nisu usko povezana s glazbom kao što su filmska i gaming industrija, odnosno bilo koja industrija u kojoj audio signali imaju konkretnu ulogu.

2.2 Edukativni segment ovog izazova

Razvijajući rješenje za ovaj izazov, imat ćete priliku istražiti i razumjeti pozadinu audio signala i njihovih digitalnih zapisa. Naučit ćete kako pripremiti audio podatke, izvući iz njih najkorisnije informacije te ih po potrebi mijenjati i prilagoditi vašem rješenju zadatka. Također imate priliku ući u dublju analizu zvuka različitih instrumenata te ljudskog glasa i istražiti njihova svojstva.

Audio klasifikacija izazovna je zbog varijabilnosti zvuka i ograničenja u dostupnim podacima za učenje. Na primjer, audio zapisi mogu sadržavati šum, propuštanja i druge nečistoće, što može otežati proces klasifikacije. Također, pojedinačne kategorije zvuka mogu biti slične jedne drugima, što može učiniti teškim njihovo razlikovanje. Stoga će polaznici, za rješavanje takvih problema, morati istražiti odgovarajuće tehnike i algoritme koji se prilagođavaju specifičnim potrebama i ciljevima klasifikacije instrumenata.

Ovaj zadatak fokusiran je na područje koje je trenutno vrlo popularno i neki segmenti još uvijek nisu sasvim svladani. Imat ćete priliku raditi sa sličnim stvarima na kojim velike tehnološke kompanije baziraju svoje najnovije projekte.

3 Zadatak

Vaš zadatak bit će razviti rješenje koje će imati mogućnost detekcije/klasifikacije instrumenata unutar audio zapisa. Dakle, za dani audio zapis (datoteka u wav formatu), potrebno je prepoznati sve instrumente koji se nalaze unutar zapisa. Mogući instrumenti su violončelo, klarinet, flauta, akustična gitara, električna gitara, orgulje, klavir, saksofon, truba, violina i ljudski pjevački glas. Oblik u kojem je potrebno vratiti rezultate modela je JSON format, pri čemu su ključevi stringovi koji predstavljaju 11 navedenih instrumenata (nazivi koje koristimo za instrumente su navedeni u opisu dataseta), a vrijednosti su 0 ili 1 ovisno o tome nalazi li se taj instrument u danom audio zapisu. Primjerice, ako se u audio zapisu javljaju samo klarinet, truba i violina (uz potencijalno neke druge instrumente koji nisu u našem popisu), ispravan output je dan sljedećim JSON-om:

```
{
  "cel": 0,
  "cla": 1,
  "flu": 0,
  "gac": 0,
  "gel": 0,
  "org": 0,
  "pia": 0,
  "sax": 0,
  "tru": 1,
  "vio": 1,
  "voi": 0
}
```

Međutim, podaci na kojima ćete temeljit vaš model bit će otežana okolnost i prilika za razvijanje različitih strategija. Naime, specifičnost ovog zadatka je u tome što se trening podaci razlikuju od onih u testnoj i validacijskoj fazi modela. Skup podataka za treniranje sadržavat će za svaki instrument određeni broj audio isječaka u kojima je on JEDINI instrument (o tome više u nastavku). Testni i validacijski podaci, s druge strane, sadržavaju audio zapise u kojima se nalazi nekoliko instrumenata, a njihova detekcija je upravo tema vašeg zadatka.

Pristup koji ćete odabrati i strategiju s kojom ćete što bolje iskoristiti trening podatke prepuštamo vama na maštu.

3.1 Opis dataset-a

Trening podaci sastojat će se od audio snimki (u trajanju od 3 sekunde) u kojima se nalazi isključivo jedan instrument. Iako, u trening podacima može doći do propuštanja još pokojeg instrumenta, npr. bubnja, ali će označen biti samo instrument koji se pojavljuje tijekom cijele dionice.

Mogući instrumenti i njihove oznake su: violončelo (cel), klarinet (cla), flauta (flu),

akustična gitara (gac), električna gitara (gel), orgulje (org), klavir (pia), saksofon (sax), truba (tru), violina (vio) i ljudski pjevački glas (voi).

Dodatno, neke od datoteka imaju bilješke u nazivu datoteke koje se odnose na prisutnost ([dru]) ili nepostojanje ([nod]) bubnjeva i glazbeni žanr: country-folk ([cou-fol]), klasika ([cla]) , pop-rock ([pop-roc]), latino-soul ([lat-sou]).

Audio datoteka: 6705 audio datoteka u 16-bitnom stereo wav formatu uzorkovanih na 44,1 kHz.

Validacijski podaci, s druge strane, sastojat će se od audio snimki(u trajanju od 5-20 sekundi) u kojima se nalazi jedan ili više instrumenta. Oznake podataka nalaze se u .txt datoteci u kojoj su navedeni svi instrumenti.

Važno je napomenuti da i u ovom slučaju u obzir dolaze samo instrumenti iz trening faze(navedeni gore), što znači da instrumenti poput bubnja neće biti navedeni.

Audio datoteka: 2874 datoteka u 16-bitnom stereo wav formatu uzorkovanih na 44,1 kHz.

Testni podaci bit će nepoznati natjecateljima i na njima će se temeljiti završna ocjena modela, odnosno njegova ukupna točnost.

4 Format rješenja

Rješenje se sastoji od dvije glavne komponente, a to su dokumentacija i programsko rješenje. Kao osnovno programsko rješenje potrebno je razviti REST API koji prihvata zvučnu datoteku i vraća predviđenu klasu. Dizajn samog API-ja i popis endpointova ostavljamo timovima na izbor, a ocjenjivat ćemo kvalitetu i jednostavnost korištenja. Osim samog API-ja, dodatno će se nagraditi i izrada bilo kakvog vizualnog sučelja u sklopu web ili mobilne aplikacije.

Sve komponente programskog rješenja potrebno je opisati u tehničkoj dokumentaciji koja predstavlja prvu komponentu dokumentacije. Očekujemo sažet i jednostavan PDF dokument koji daje upute za pokretanje i korištenje vaše aplikacije, a jednostavnost pokretanja aplikacije utjecat će na konačnu ocjenu. Datoteke potrebne za korištenje aplikacije predavat će se kao zasebna komponenta rješenja, a sam format ostavljamo natjecateljima na izbor. Preporučamo exportati Conda environment uz kod ili Docker kontejner.

Osim tehničke dokumentacije, potrebno je predati i projektnu dokumentaciju. To je duži dokument koji potpuno opisuje vaše rješenje, počevši od analize podataka pa sve do konačnog modela. Poželjno je dokumentaciju potkrijepiti vizualizacijama podataka i rezultata koje olakšavaju razumijevanje. Uz to, bodovat će se jasnoća i razumljivost same dokumentacije.

Na kraju natjecanja, evaluirat ćemo vaše rješenje na konačnom testnom skupu podataka. Rezultati na tom skupu podataka predstavljat će jednu komponentu za ključne ocjene.

Finalno, natjecatelji šalju konačno rješenje u obliku zip arhive koja sadrži

- Projektnu dokumentaciju u PDF formatu
- Tehničku dokumentaciju u PDF formatu
- Izvorni kod rješenja
- Kod i ostale potrebne elemente za podizanje aplikacije

Najbolji timovi nakon prvog dijela natjecanja bit će pozvani na završnicu za koju će biti potrebno pripremiti prezentaciju u trajanju od 15 minuta.

5 Elementi ocjenjivanja

Kriteriji	Raspon ocjena	Doprinos u ocjeni	Bilješka
Projektna dokumentacija	0-5	15%	Ocjenjuje se ukupni doživljaj dokumentacije. Općenito razumijevanje problema i implementiranog rješenja te način na koji je to prikazano i objašnjeno. To uključuje i razne vizuale, dijagrame i primjere.
Tehnička kvaliteta	0-5	20%	Ocjenjuje se sama tehnička izvedba (struktura i organizacija koda) rješenja te uz to priložena tehnička dokumentacija.
Kvaliteta rješenja	0-5	25%	Ocjenjuje se pristup i način realizacije rješenja te njegova kreativnost, omjer kvalitete i veličine i robusnost na dodavanje nekih novih funkcionalnosti ili manjih promjena
Preciznost modela	0-5	15%	Za svaki audio zapis testnog skupa podataka, izračunat ćemo prosječan broj pogodenih instrumenata i zatim uprosječiti dobivene brojeve po ukupnom broju audio zapisa. Taj broj bit će jednak ukupnoj točnosti modela.
Prezentacija	0-5	20%	Ocjena prezentacija ovisi o vizualnom aspektu slajdova, kolika količina bitnih informacija je prezentirano, jasnoća prezentacije te sveukupni performans.
Predaja rješenja	0-5	5%	Od ekipa se minimalno očekuje da imaju samo API na kojem će se testirati rješenje, međutim bilo kakav dodatno uloženi trud u korisničko sučelje (web ili mobilna aplikacija) biti će dodatno nagrađeno