

Prepoznavanje ugođaja pjesme

Marija Gegić¹ i David Mrkoci¹

¹ Sveučilište u Zagrebu, Prirodoslovno matematički fakultet

Prepoznavanje ugođaja pjesme je kompleksan problem kod kojeg se općenite pjesme pokušavaju kategorizirati u neke unaprijed određene kategorije. U ovom radu iznosimo opis tog problema te predlažemo jedan mogući pristup rješavanju korištenjem metoda strojnog učenja. Taj smo pristup implementirali i testirali na odabranom setu pjesama, koje smo dijelili u 4 moguće kategorije: sretne, tužne, energične, mirne.

1 Uvod

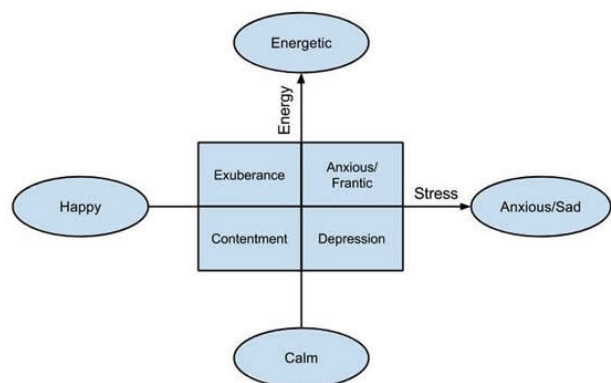
U razdoblju 21. stoljeća svake godine se objavi više od 130 milijuna novih pjesama. S obzirom na toliku veliku količinu podataka, prirodno se pojavljuje potreba za automatiziranim načinom obrade i analize podataka vezanih uz te pjesme. Prepoznavanje ugođaja pjesme je vrlo značajan i zanimljiv problem u tom području.

Preciznim prepoznavanjem ugođaja pjesme možemo uvelike poboljšati razne postojeće alate, aplikacije i servise za slušanje glazbe i kreiranje playlisti, jer bi nam to pomoglo da stvaramo bolje prijedloge novih pjesama za korisnike. To je vrlo korisno i primjenjivo u današnje doba, kad se velik broj online servisa trudi personalizirati sadržaj za svoje korisnike i unaprijediti svoje sustave za preporuke. Korisnici bi tako mogli jednostavnije pretraživati pjesme ovisno o ugođaju i kategoriji u koju spadaju. Izvan toga, prepoznavanje ugođaja je korisno i u filmskoj i medijskoj industriji, primjerice za traženje dobrih pjesama koje ugođajem odgovaraju nekom filmu, reklamama, video sadržajima i slično.

2 Opis problema

U sklopu ovog projekta bavili smo se razvijanjem modela za prepoznavanje ugođaja pjesme. Proučavali smo

4 moguća ugođaja i pjesme kategorizirali kao sretne ili tužne, te mirne ili energične. Ove 4 značajne kategorije smo odabrali kao podskup klasifikacijskih kategorija koje je predložio psiholog Robert Thayer u svom modelu raspoloženja [2], kao što je prikazano na donjoj slici.



Slika 1: Model raspoloženja

Pri tome se ograničavamo na pjesme koje su objavljene na servisu Spotify. Korištenjem Spotify API-ja možemo izvući razne informacije o pjesmama, poput izvođača, albuma, godine izdavanja i određenih značajki objavljenog zvučnog zapisa te pjesme. Nama su najzanimljivije bile upravo te značajke, jer smo njih koristili kao glavni dio dataseta koji nam služi za treniranje i testiranje modela. Nakon razvijanja modela, moći ćemo kategorizirati proizvoljnu pjesmu dostupnu na Spotifyju, koristeći značajke njezinog zvučnog zapisa. Podatke koje smo koristili za razvijanje modela, kao i značajke koje su dostupne putem Spotify API-ja, predstaviti ćemo u daljnjim odlomcima.

3 Pregled dosadašnjih istraživanja

Pri pregledu dosadašnjih istraživanja ističemo [2], [1] kojima su postavljene osnovne kategorizacije koje koristimo u ovom projektu. Sličan je problem obrađen u članku [3], gdje je autor razvio model korištenjem 800 pjesama, koje je kategorizirao u jednu od 4 kategorije: sretna, tužna, mirna, energična. Naš pristup donosi nekoliko vrsta unaprijeđenja pri usporedbi s metodom korištenom u tom članku. Za početak, odlučili smo koristiti veći i noviji dataset, jer smatramo da korištenje samo 800 pjesama može polučiti nedovoljno precizne rezultate. Također, odlučili smo razvijati dvije vrste modela - poseban model za prepoznavanje sret-nih/tužnih pjesama, te poseban model za prepoznavanje energičnih/mirnih pjesama.

4 Podaci

Pjesme za treniranje i testiranje modela smo prikupili ručnim pretragama playlisti dostupnih na Spotifyju. Odabrali smo playliste koje sadrže sretne, tužne, mirne ili energične pjesme, pri čemu smo obratili posebnu pozornost na to da biramo playliste koje ugođaj imaju nedvojbeno jasno naveden u svom nazivu, te playliste koje imaju velik broj *likeova*, koji impliciraju da se korisnici Spotifyja slažu s kategorizacijom. To znači da su svaku kategorizaciju pjesama potvrdile desetine tisuća ljudi. Sljedeća tablica prikazuje veličinu dataseta, tj. broj pjesama za svaki ugođaj.

kategorija	broj pjesama
energične	1391
mirne	1345
sretne	1195
tužne	812

Tablica 1: Pjesme u datasetu

Pjesme smo zatim podijelili u 2 dataseta: happy/sad dataset i calm/energetic dataset, te razvijali modele za svaki posebno.

Kao što smo prethodno naveli, podatke o pjesmama smo dobili korištenjem Spotify API-ja. Sljedeća tablica prikazuje jedan primjer podataka koje pamtimo za pojedinačnu pjesmu. U primjeru se radi o pjesmi *We Will Rock You* grupe *Queen*. Pri tome su svi podaci prikazani u tablici dobiveni kao rezultat queryja na Spotify API, osim zadnjeg retka, koji prikazuje ugođaj pjesme. Tu smo informaciju sami postavljali prilikom prikupljanja podataka, ovisno o tome u kakvoj vrsti playliste se pjesma nalazi. Konkretno, ova se pjesma nalazila u nekoj playlisti koja sadrži energične pjesme, stoga smo postavili vrijednost *mood* na *'energetic'*, a ostale podatke smo dobili Spotify API queryem. Ova pjesma doista jest energična, pa je jasno da je dobro kategorizirana.

name	We Will Rock You
album	News Of The World
artist	Queen
id	4pbJqGIASGPr0ZpGpnWkDn
release_date	1977-10-28
popularity	79
length	122066
danceability	0.692
acousticness	0.676
energy	0.497
liveness	0.259
valence	0.475
loudness	-7.316
speechiness	0.119
instrumentalness	0.00
tempo	81.308
key	2
time_signature	4
mood	energetic

Tablica 2: Primjer prikupljenih podataka za jednu pjesmu

5 Odabir značajki

Nakon pregleda svih podataka koje smo mogli prikupiti kroz Spotify API, morali smo odabrati koje podatke želimo koristiti za treniranje modela. Istražili smo njihova značenja u teoriji glazbe i u Spotifyjevoj analizi, i zaključili smo da su nam najbitnije značajke za početak: *length*, *danceability*, *acousticness*, *energy*, *instrumentalness*, *liveness*, *valence*, *loudness*, *speechiness*, *tempo*. Nismo htjeli koristiti podatke o popularnosti, starosti i izvođaču, jer smo htjeli dobiti model koji će se moći koristiti i za pjesme koje su u potpunosti novoizdane (time bi model bio pogodan za primjene koje smo prethodno naveli, poput preporuke novih pjesama). Bez tih značajki, stvorit ćemo model koji ovisi isključivo o zvučnom zapisu pjesme, a ne o nekim dodatnim informacijama koje postoje izvan samog zvučnog zapisa.

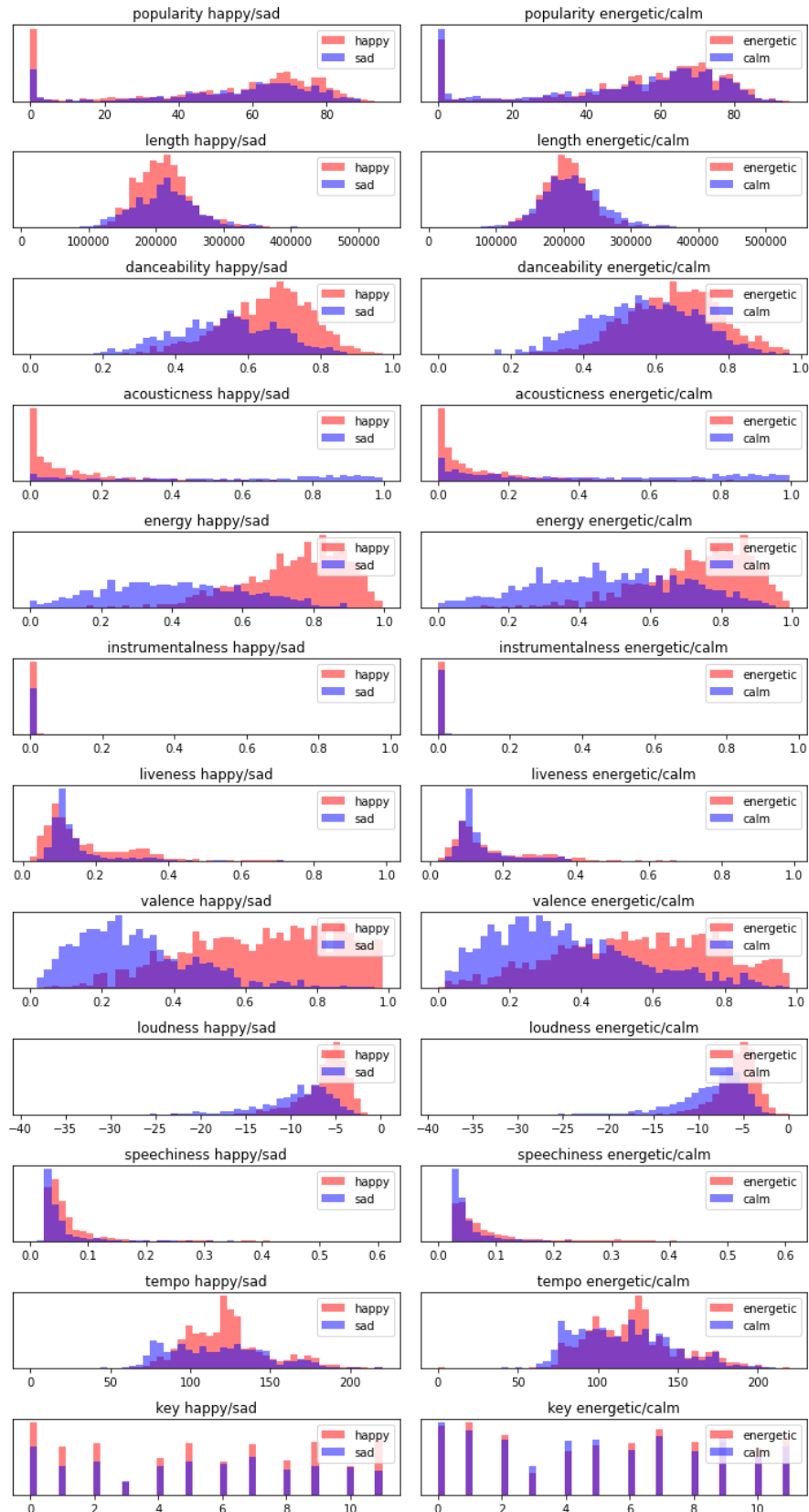
Dodatno, odlučili smo ne koristiti vrijednosti *instrumentalness* jer smo se pri odabiru pjesama ograničili na vokalne pjesme, za koje je ta vrijednost bila jednaka ili gotovo jednaka 0. Eksperimentalno smo i provjerili da ta značajka ne bi bila relevantna, usporedbom rezultata kad ona je/nije korištena pri treniranju.

6 Obrada podataka i smanjivanje dimenzija

Kao što je bilo vidljivo u prethodnoj tablici, podaci koje dobivamo od Spotify API-ja nisu normalizirani, stoga ih je bilo potrebno normalizirati. Za to smo koristili *MinMaxScaler()*. Nadalje, odlučili smo istražiti ima li potrebe za smanjivanjem dimenzionalnosti, jer smo baratali većim brojem značajki, od kojih bi neke mogle biti korelirane. Kao što smo već napomenuli, primijetili

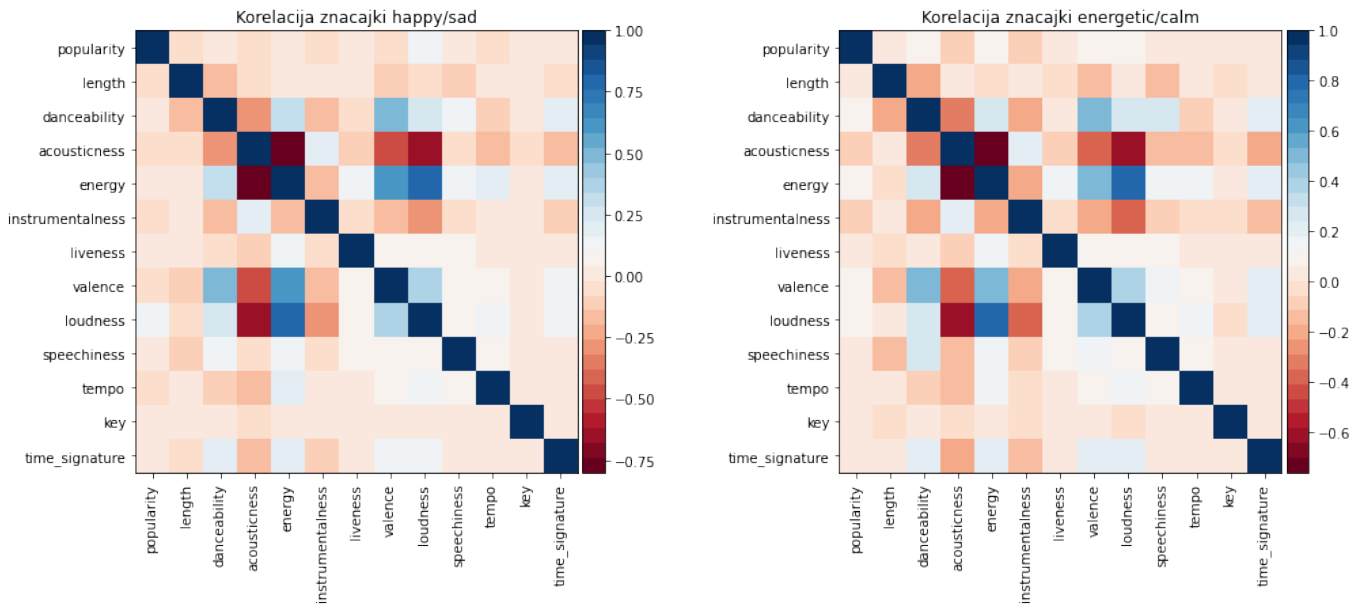
smo da značajka *instrumentalness* poprima vrijednost 0 za jako velik broj pjesama, te da ju možemo zanemariti. Sljedeća slika grafički prikazuje raspodjelu podataka

koje smo prikupili, prije normalizacije korištenjem *MinMaxScaler()*.



Slika 2: Raspodjela vrijednosti značajki

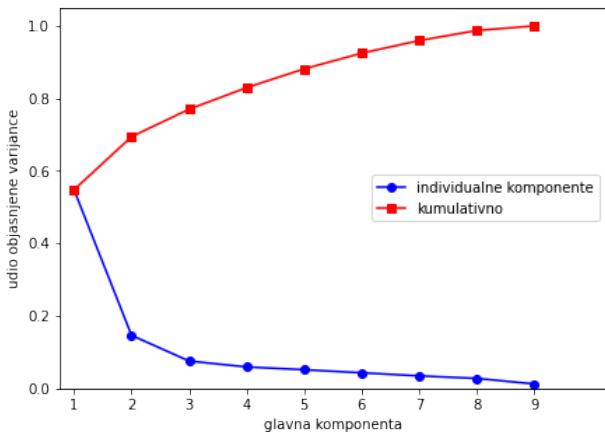
Proučavali smo i korelaciju između značajki. Sljedeća slika prikazuje korelaciju značajki kod pjesama za happy/sad i energetic/calm datasetove.



Slika 3: Korelacija značajki

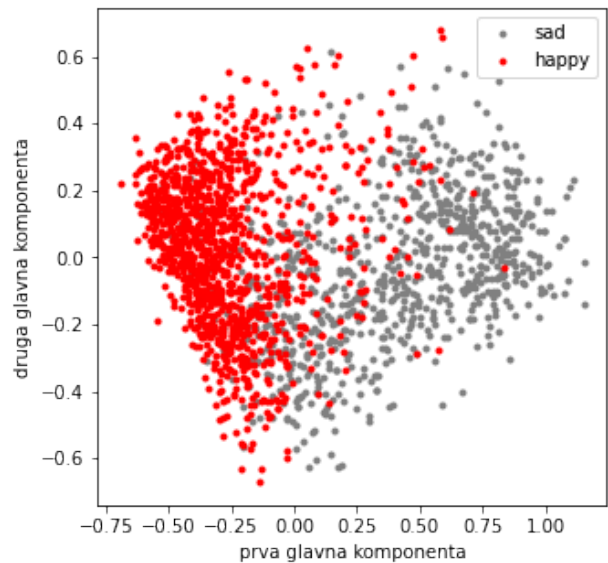
Proveli smo postupak analize glavnih komponenti kako bismo bolje shvatili povezanost značajki. Grafički ćemo prikazati iznos varijance i kumulativne varijance nakon provođenja PCA, kao i projekciju podataka na prve dvije glavne komponente za oba dataseta.

Za početak, promotrimo varijancu objašnjenu sa svakom od glavnih komponenti, kao i kumulativnu varijancu objašnjenu uključivanjem svake naredne glavne komponente u našem projiciranom datasetu.



Slika 4: Varijance za happy/sad dataset

Primjećujemo da prve dvije komponente zajedno objašnjavaju više od 70% varijance u podacima. Pogledajmo kako naši podaci izgledaju kada se projiciraju na prve dvije glavne komponente.

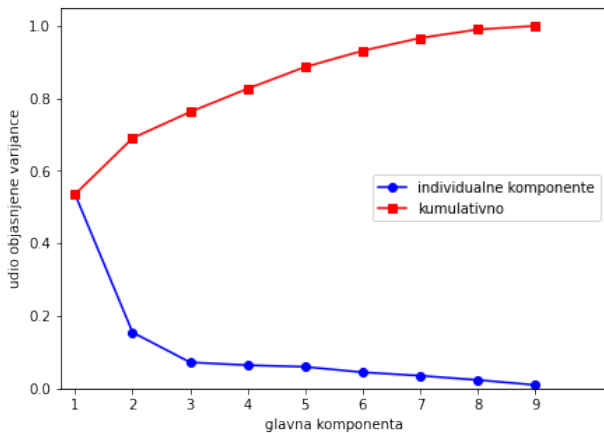


Slika 5: Projekcija podataka za happy/sad dataset

Vidljivo je da sad imamo puno značajniju razliku između sretnih i tužnih pjesama. Na gornjem grafu se posebno ističe gust cluster točaka za sretne pjesme.

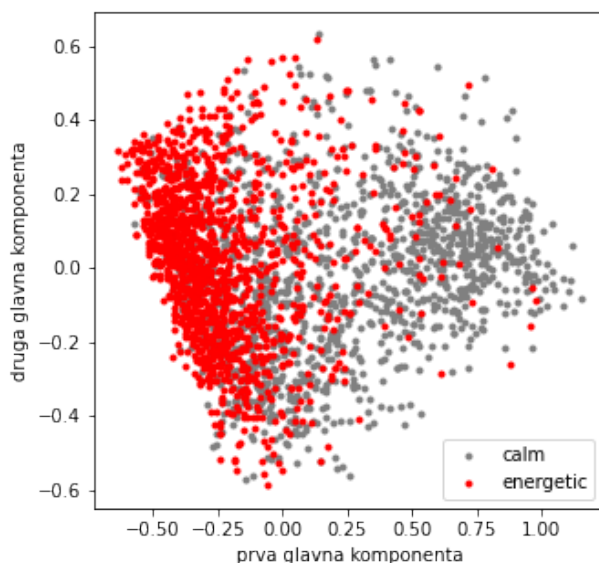
Isti postupak smo proveli i za dataset mirnih i energičnih pjesama.

Ponovno promotrimo varijancu objašnjenu sa svakom od glavnih komponenti, kao i kumulativnu varijancu objašnjenu uključivanjem svake naredne glavne komponente u našem projiciranom datasetu.



Slika 6: Varijance za calm/energetic dataset

Dobiveni graf je veoma sličan onom u prethodnom slučaju. Ponovno prve dvije komponente zajedno objašnjavaju više od 70% varijance u podacima. Sljedeća slika prikazuje kako podaci izgledaju kada se projiciraju na prve dvije glavne komponente.



Slika 7: Projekcija podataka za calm/energetic dataset

Ponovno vidimo unapređenje razlike među podacima, te se ponovno pojavljuje jedan gust cluster točaka, u ovom slučaju za energetske pjesme.

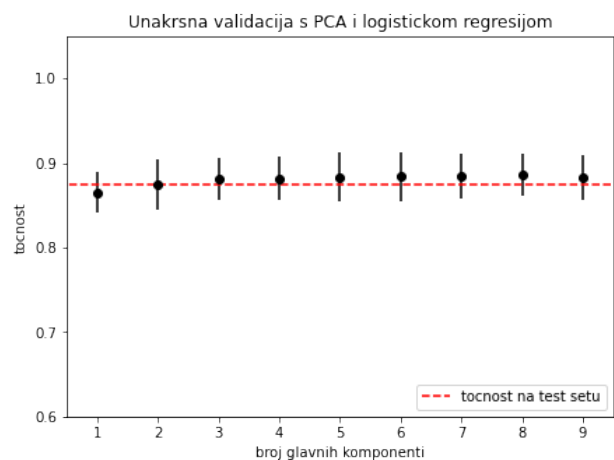
7 Modeliranje

Nakon primjene PCA projekcije za smanjenje dimenzija i dekokoreliranje značajki, možemo započeti s kreiranjem modela. Kao tehniku klasifikacije

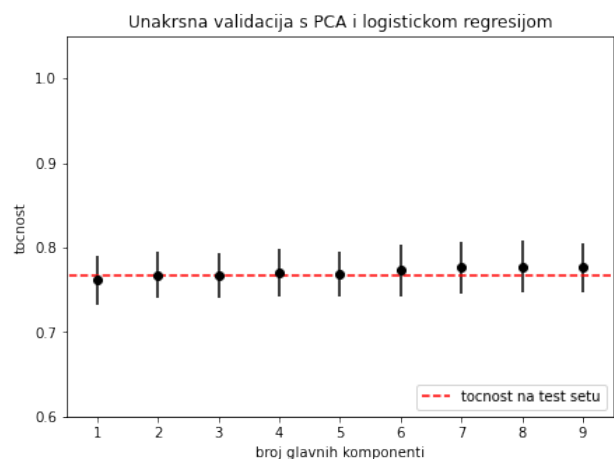
koristimo logističku regresiju implementiranu u `sklearn.linear_model.LogisticRegression`. Točnost naših predikcija evaluiramo unakrsnom validacijom implementiranom u `sklearn.grid_search.GridSearchCV` za različiti broj glavnih komponenti. Na kraju ćemo najbolje modele evaluirati na zasebnom test skupu.

Ovo činimo za svaki dataset zasebno, odnosno razvijamo dva modela neovisna jedan o drugom. Od početnih datasetova, koristimo 30% podataka za testiranje, a 70% za treniranje modela, kreiranih funkcijom `train_test_split`.

Sljedeće dvije slike prikazuju točnost na testiranim setovima za unakrsnu validaciju s PCA i logističkom regresijom.



Slika 8: Točnost za happy/sad dataset

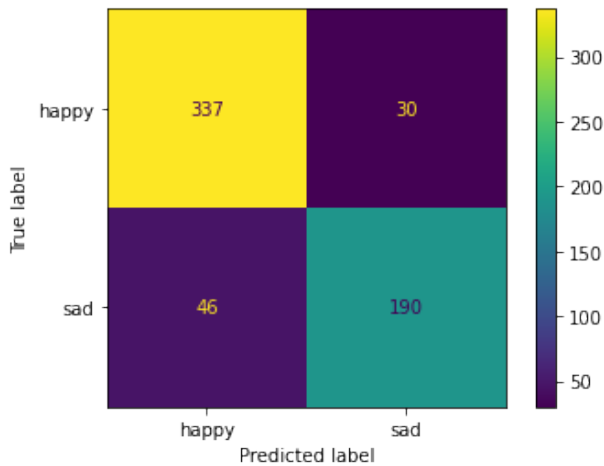


Slika 9: Točnost za calm/energetic dataset

Vidljivo je da je postignuta bolja točnost kod prvog modela, što ćemo dodatno komentirati u sljedećim odlomcima.

8 Rezultati

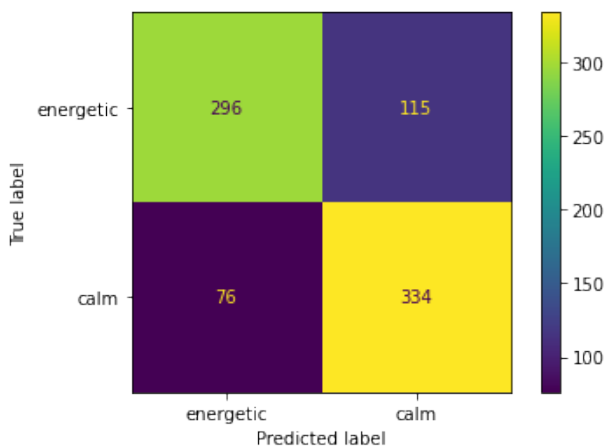
Sljedeća slika prikazuje konfuzijsku matricu za model za sretne/tužne pjesme, testiran na testnom setu koji sadrži 30% pjesama iz happy/sad dataseta.



Slika 10: Konfuzijska matrica za happy/sad dataset

Točnost modela je 87.4%.

Sljedeća slika prikazuje konfuzijsku matricu za model za mirne/energične pjesme, testiran na testnom setu koji sadrži 30% pjesama iz calm/energetic dataseta.



Slika 10: Konfuzijska matrica za calm/energetic dataset

Točnost modela je 76.6%.

9 Diskusija

Usporedimo li naše dobivene rezultate s rezultatima dobivenim u [3], vidimo da smo ostvarili bolju točnost modela, pogotovo u slučaju sretnih i tužnih pjesama.

Gledajući matrice konfuzije, vidimo da je model za prepoznavanje sretnih i tužnih pjesama dosta precizan. Bolji je u prepoznavanju sretnih pjesama nego tužnih, redom s preciznostima od 88% i 86.3% za sretne i tužne pjesme. Model za prepoznavanje mirnih i energičnih pjesama je manje precizan, pri čemu malo bolje prepoznaje energične pjesme nego što prepoznaje mirne pjesme, redom s preciznostima 83.9% za energične i 74.4% za mirne pjesme.

Kao moguće objašnjenje za slabiju preciznost prepoznavanja mirnih pjesama, prokomentirat ćemo strukturu vokalnih pjesama. Najčešći format moderne popularne glazbe sastoji se od uvodnog dijela (uvod, intro), stihova, pre-refrena, refrena, mostova i završnog dijela (outro), pri čemu se stihovi, pre-refren, mostovi i refren mogu pojavljivati i više puta. Intro i outro pjesama su najčešće mirniji i ne sadrže tekst, a intenzitet pjesme jača prema refrenu. Očekivano je da će postojati mirniji dijelovi pjesme čak i kod energičnih pjesama, zbog čega je bilo teže izraditi model za energične/mirne pjesme nego što je to slučaj kod sretnih/tužnih pjesama.

Potencijalni način za bolju analizu pjesama i dobivanje boljih modela mogao bi se dobiti proučavanjem pjesama u dijelovima - zasebnim analiziranjem srednjeg dijela pjesme, te intro i outro dijela. Taj korak izlazi iz okvira ovog projekta, jer smo se ograničili na proučavanje podataka dostupnih kroz Spotify, gdje je postojala samo analiza cijele pjesme.

Literatura

- [1] *Machine Learning Approaches for Mood Classification of Songs toward Music Search Engine*. URL: <https://ieeexplore.ieee.org/document/5361715>.
- [2] *Music Mood Classification*. URL: <https://sites.tufts.edu/eeseniordesignhandbook/2015/music-mood-classification/>.
- [3] *Predicting the Music Mood of a Song with Deep Learning*. URL: <https://towardsdatascience.com/predicting-the-music-mood-of-a-song-with-deep-learning-c3ac2b45229e>.