

# Uvrščanje glasbe

Matic Knez

## Podatki

- viri podatkov (developer platforma spotify – knjižnica Spotipy)
- prvotni namen zbiranja (na voljo veliko podatkov, fleksibilnost podatkov)
- tip podatkov (naslovi, izvajalci, lastnosti, datum, album...)
- obseg podatkov (80.000+, 15 atributov...)
- število manjkajočih zapisov, napak, druge težave (žanri, datumi, duplikati)
- opis predprocesiranja, da postanejo podatki “uporabni” (odstranitev duplikatov)

## Glavna vprašanja/cilji podatkovnega rudarjenja

- Uvrščanje glasbe glede na vzdušje
- Uvrščanje glasbe glede na žanr
- Uvrščanje glasbe na playlist

# Audio lastnosti pesmi

Danceability - kako primerna je skladba za ples

Acousticness - ali je skladba akustična

Energy - kako intenzivna in aktivna je skladbe.

Instrumentalness - ali skladba vsebuje vokale.

Liveness - ali je prisotnot publike (nastop v živo).

Loudness - povprečna glasnot skladbe v decibelih(dB

Speechiness - kako verjetno je, da je govor v skladbi.

Tempo - BPM (Beats per minute)

Valence - kako pozitivna je skladba.

Mode - Tonaliteta (dur ali mol) skladbe

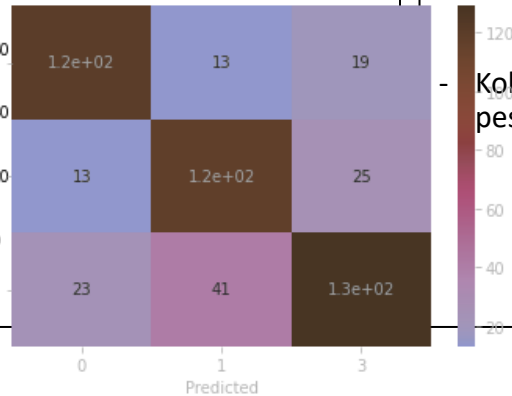
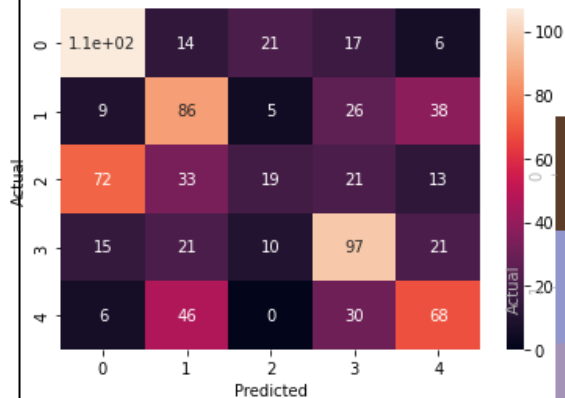
Key - V katerem ključu je skladba

Duration\_ms - Dolžna pesmi v mili sekundah.

## Podroben opis ciljev in metod

### - Uvrščanje glasbe glede na vzdušje

- 5 vzdušji (za vsako enako število podatkov)
- 6 klasifikacijskih modelov
  - Logistic Regression
  - Support Vector Classification
  - Random Forest
  - Multi-layer Perceptron (Neural Network)
  - K-Nearest Neighbours
  - Decision Tree



## Rezultati/dosedanje ugotovitve/odprta vprašanja

### Analiza podatkov

- Lastnosti pri žalostno in mirno ter veselo in živahno so si zelo podobne.

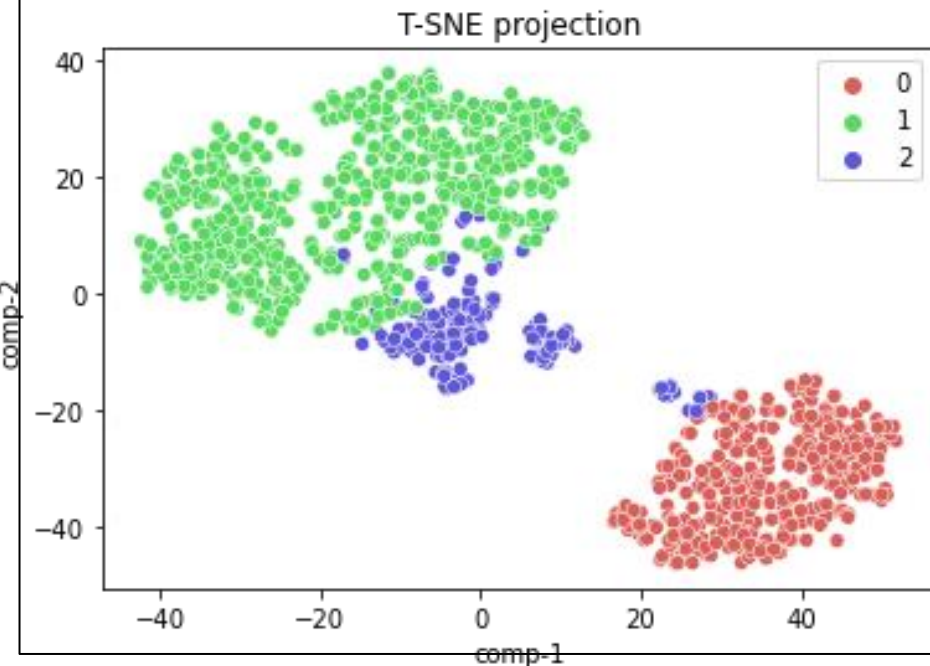
### Rezultati

- Najbolje se izkaže Random Forest, s točnostjo okoli 0.46.
- Veliko pesmi označenih z mirno klasificira kot žalostne. Kar nekajkrat pa zameša tudi vesele in živahne pesmi.
- Iste klasifikacijske modele poiskujemo še na samo 3 počutjih, ki so si med seboj različna.
- Tokrat so rezultati boljši. Random Forst doseže točnost okoli 0.73.
- Koliko bi se rezultati izboljšali, če bi dodali še analizo besedila pesmi?

## Podroben opis ciljev in metod

### - Združevanje glasbe v skupine

- K-Means Clustering
  - Elbow metoda za izračun optimalne vrednosti k.



## Rezultati/dosedanje ugotovitve/odprta vprašanja

### Rezultati

- Skupine so med seboj najbolj jasno ločene, če je k enako 3.
- Po analizi skupin lahko razberemo, da se 2. skupina od ostalih dveh zelo razlikuje še posebej po 'acousticness', 'energy', 'loudness' in 'valence'. V tej skupini so pesmi bolj mirne in akustične.
- Ostali dve skupini pa sta si medseboj zelo podobni in sta bolj živahni in polni energije.

## 2. Uvrščanje glasbe v žanr

- Classical
- Rap
- Rock
- EDM
- Latin
- pop

Vseh pesmi v data setu: 83698

pop : 11309

rock : 18330

rap : 11365

edm : 8693

latin : 14129

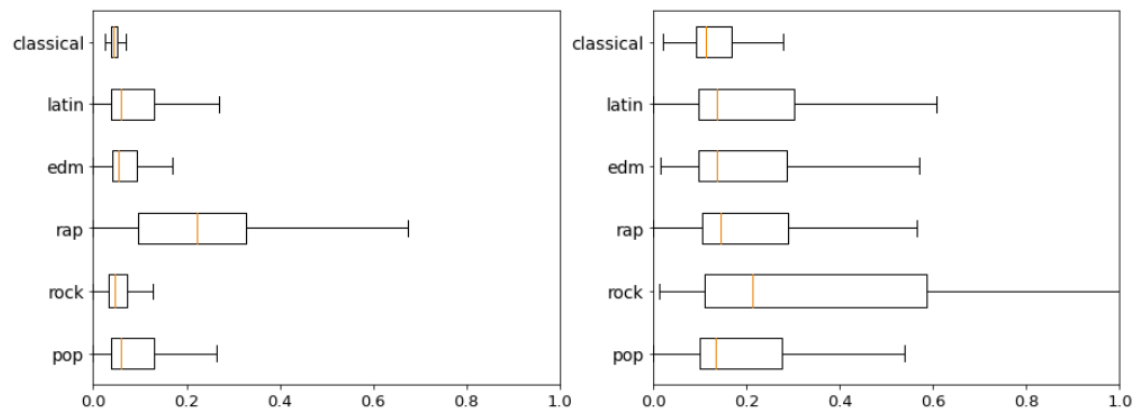
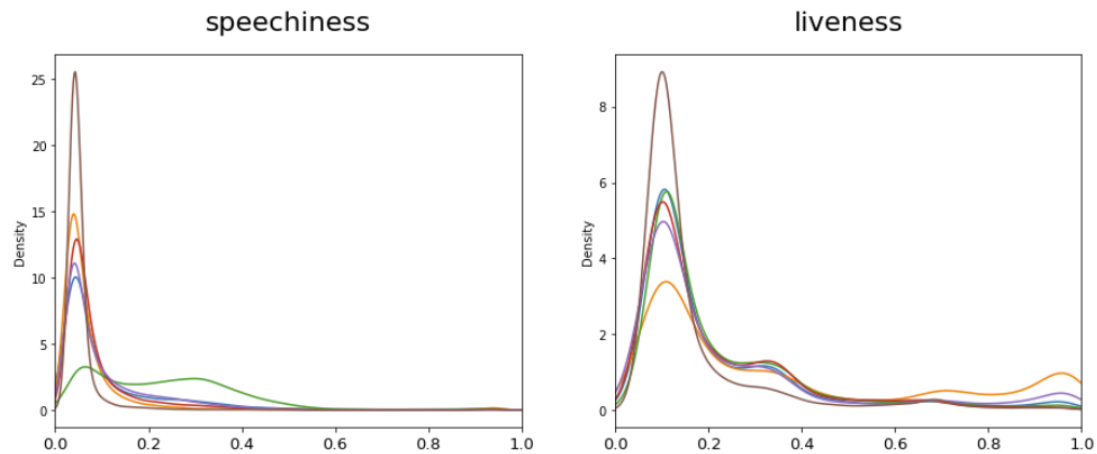
classical : 19872

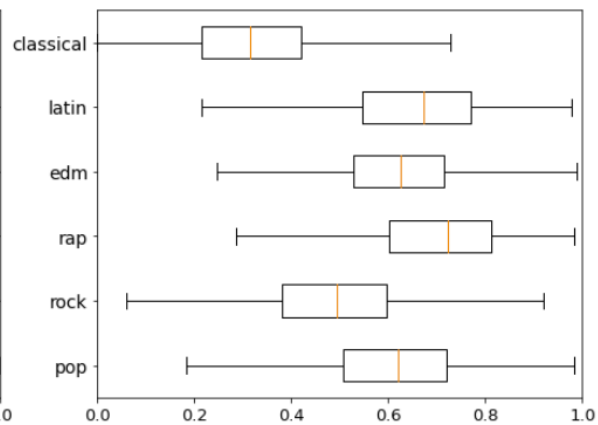
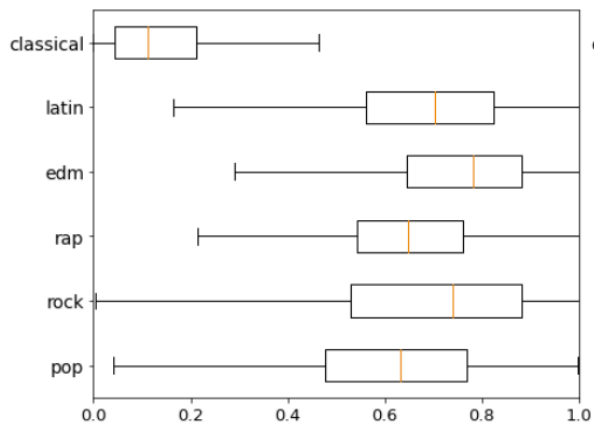
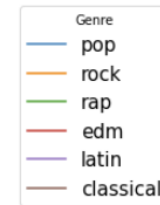
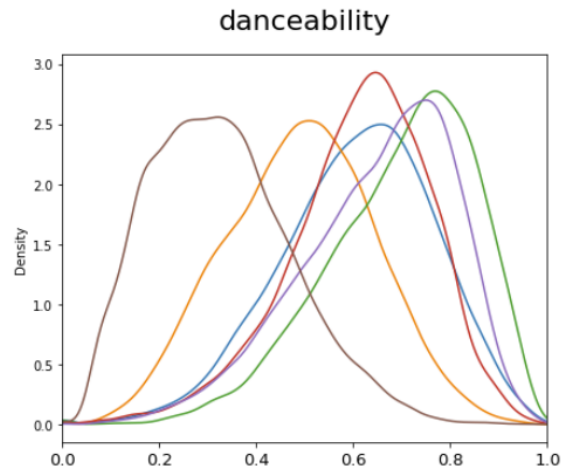
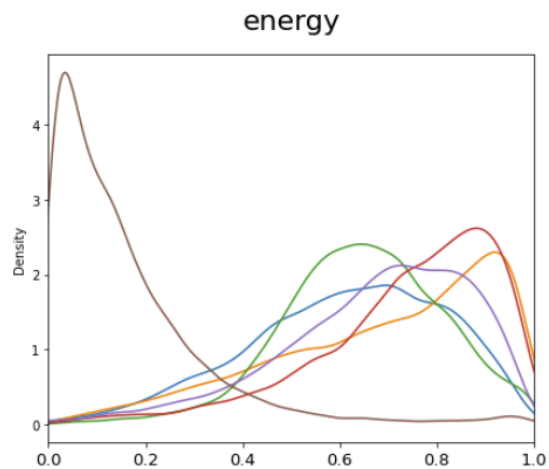
Tracks.csv

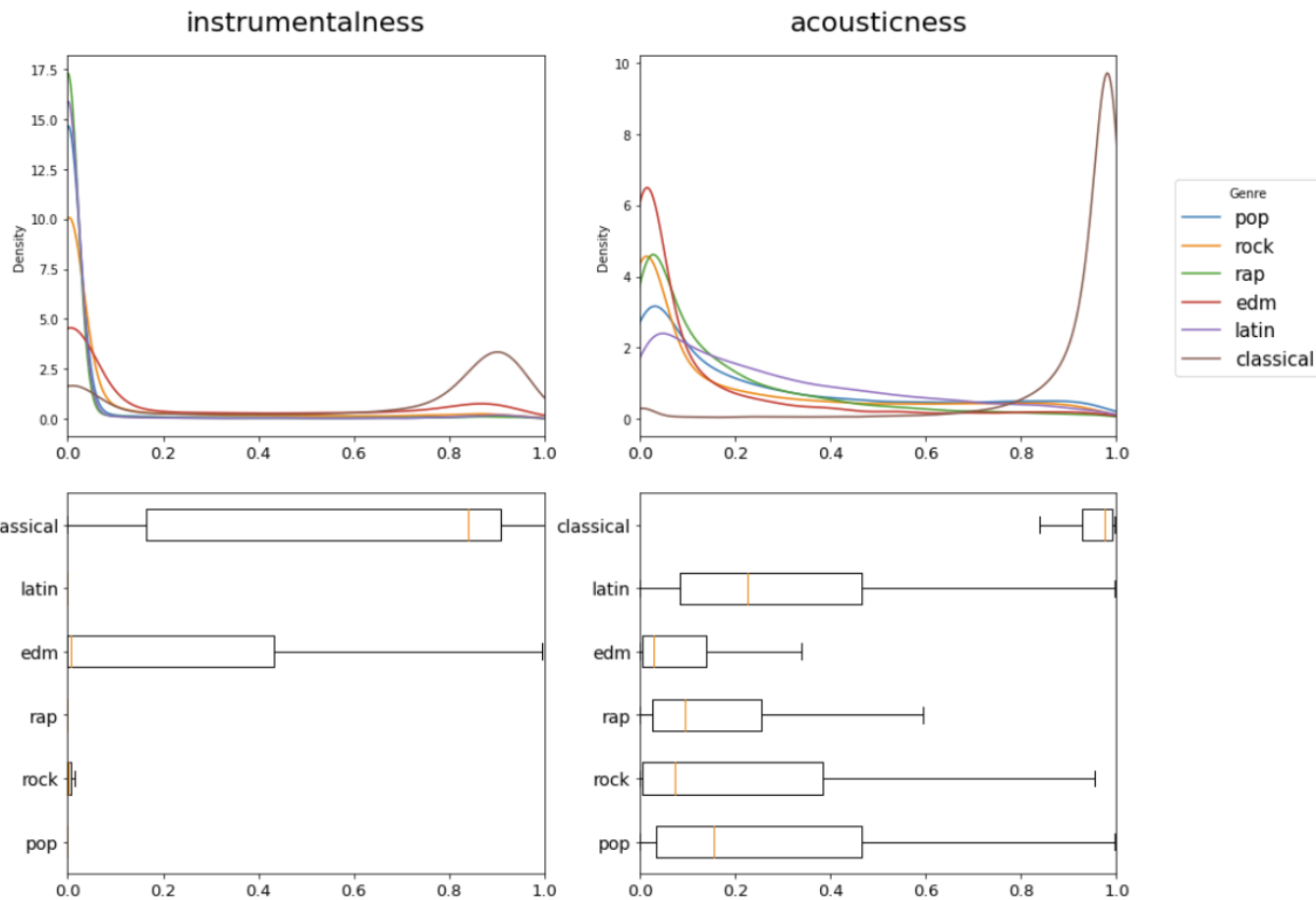
id	name	artist_id	genre	album_id
2HSmyk2qMN8WQJuGhaQgCk	Champagne Poetry	3TVXtAsR1Inumwj472S9r4	rap	3SpBlxme9WbeQdI9kx7KAV
6jy8yJfgCsMHdu2Oz4BGKX	Papi's Home	3TVXtAsR1Inumwj472S9r4	rap	3SpBlxme9WbeQdI9kx7KAV
37Nqx7iavZpotJSDXZWbJ3	Girls Want Girls (with Lil Baby)	3TVXtAsR1Inumwj472S9r4	rap	3SpBlxme9WbeQdI9kx7KAV
61S79KIVA4l9FXbnsylEHT	In The Bible (with Lil Durk & Giveon)	3TVXtAsR1Inumwj472S9r4	rap	3SpBlxme9WbeQdI9kx7KAV
4VCbgldr8ptegWeJpqLVHH	Love All (with JAY-Z)	3TVXtAsR1Inumwj472S9r4	rap	3SpBlxme9WbeQdI9kx7KAV

Audio\_Features.csv

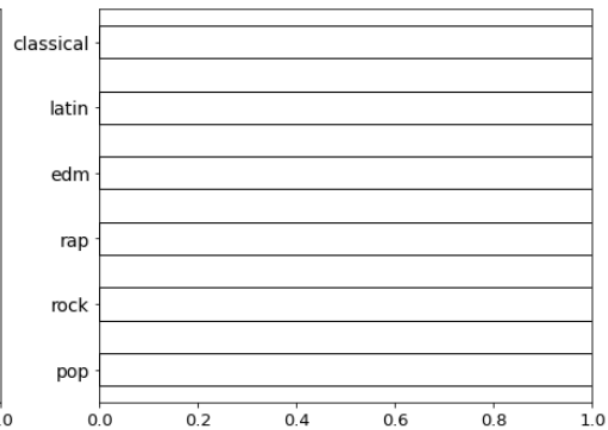
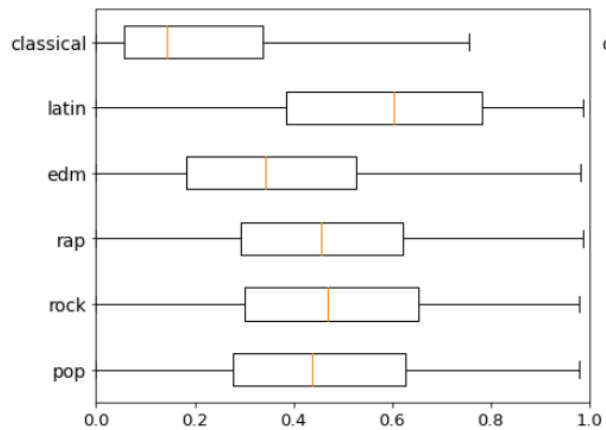
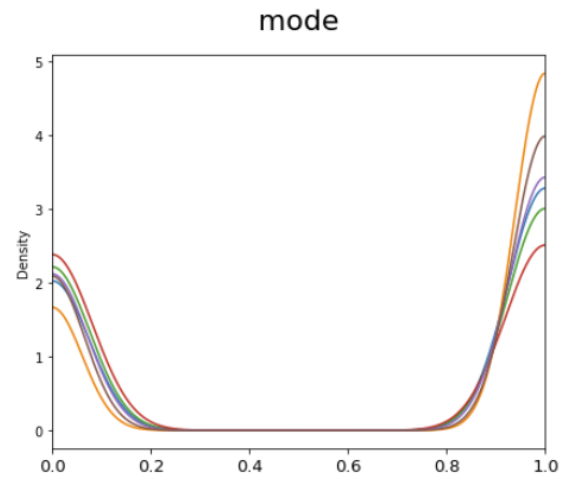
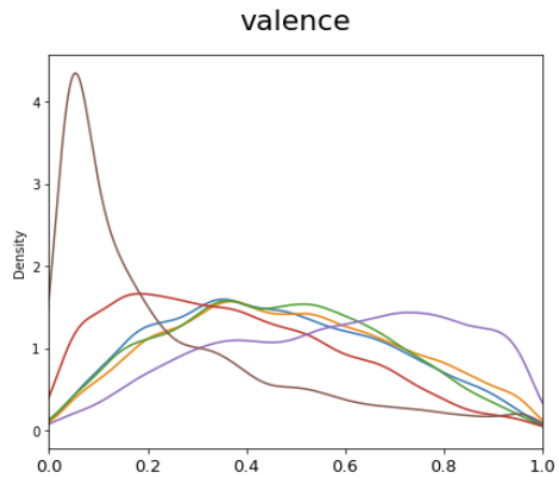
danceability	energy	key	loudness	mode	speechiness	acousticness	instrumentalness	liveness	valence	tempo	duration_ms
0.423	0.667	6	-7.012	0	0.326	0.758	0.000000	0.246	0.496	86.743	336511
0.701	0.741	2	-6.157	1	0.313	0.112	0.000000	0.699	0.588	140.177	178623
0.767	0.438	7	-8.726	0	0.290	0.181	0.000000	0.142	0.381	86.975	221980
0.686	0.385	7	-8.350	0	0.297	0.614	0.000000	0.131	0.147	143.070	296568
0.730	0.576	2	-5.442	1	0.287	0.354	0.000036	0.150	0.155	92.131	228461

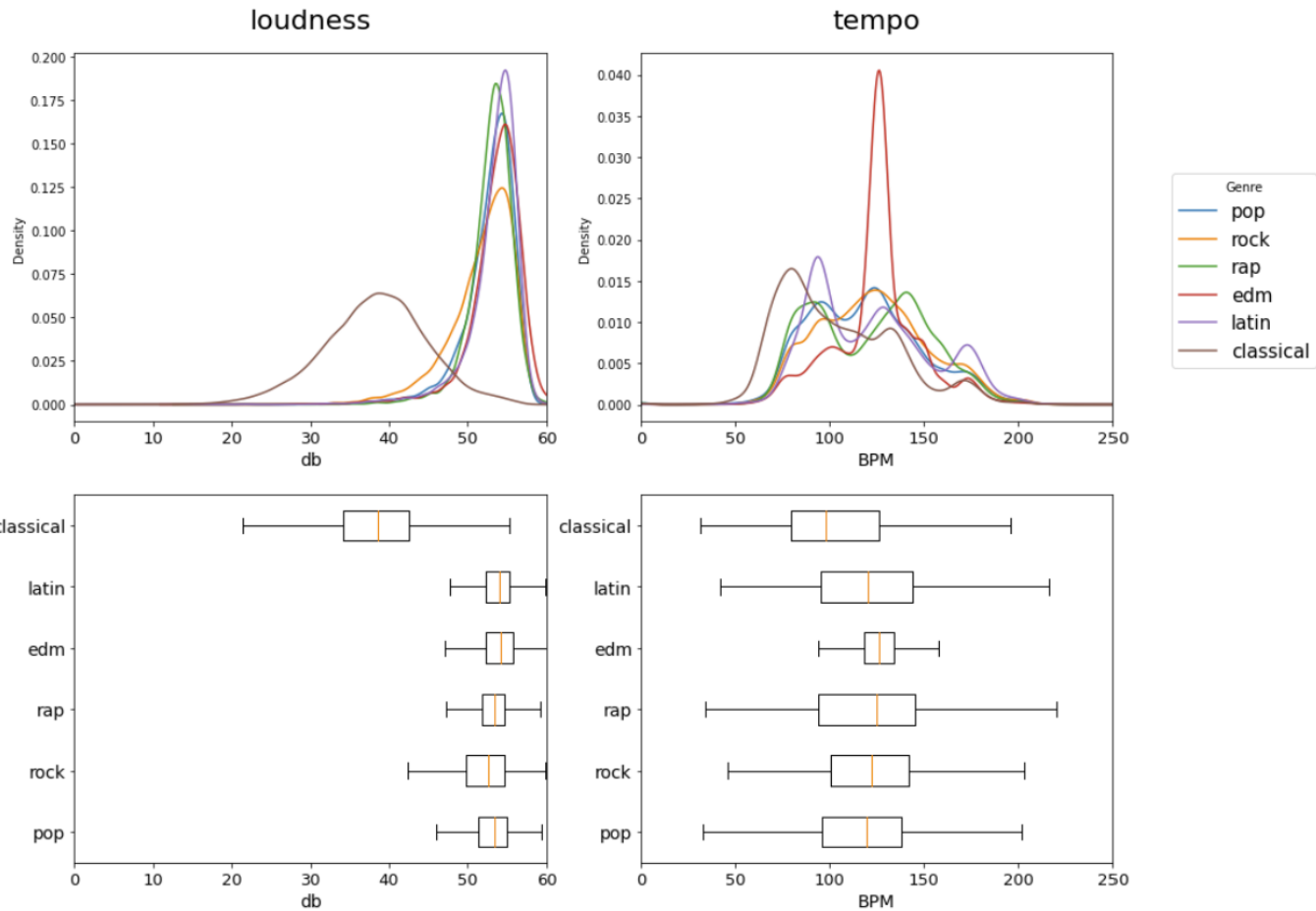












# Osamelci

## Prej

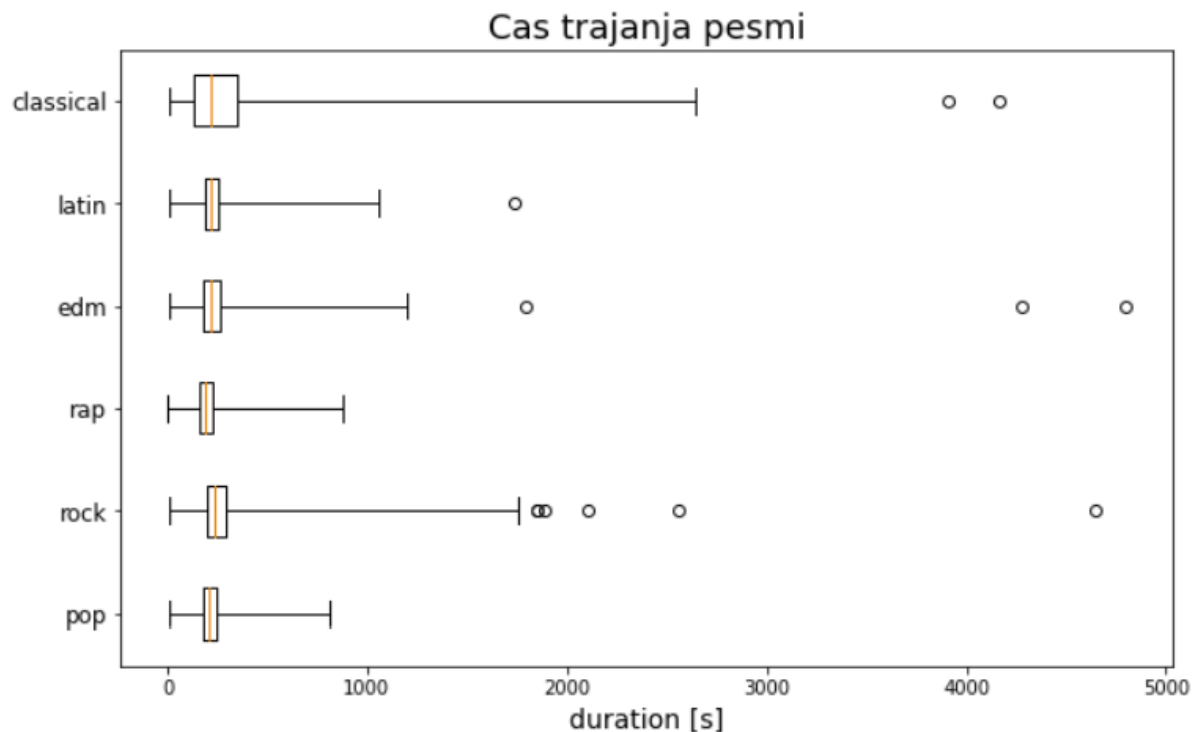
Max dolzina: 4794399ms (4794s) (79 min)

Min dolzina: 3677ms (3s) (0 min)

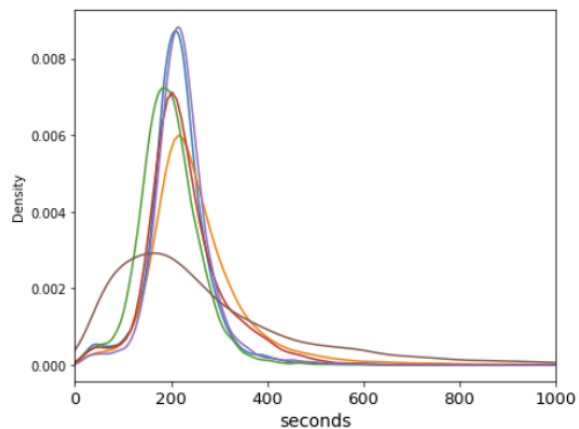
## Potem

Max dolzina: 1279467ms (1279s) (21 min)

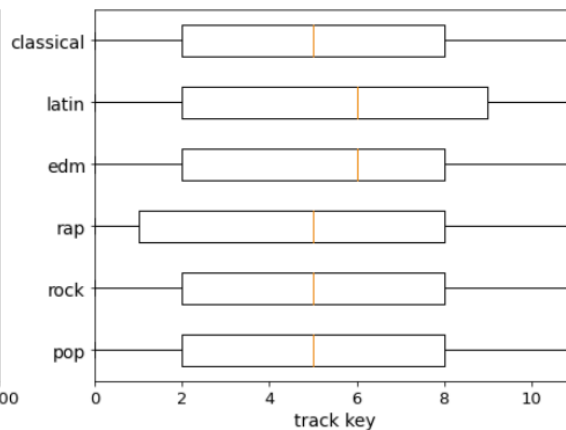
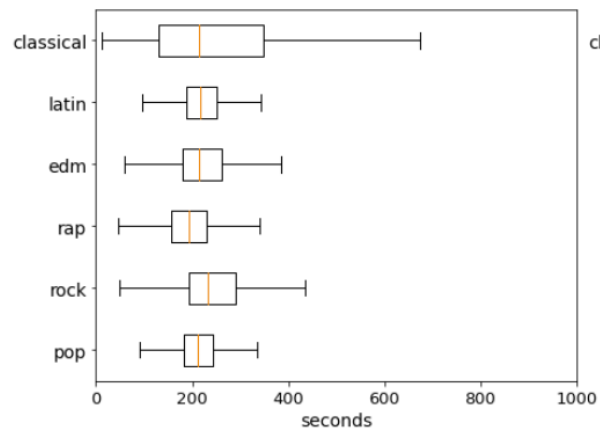
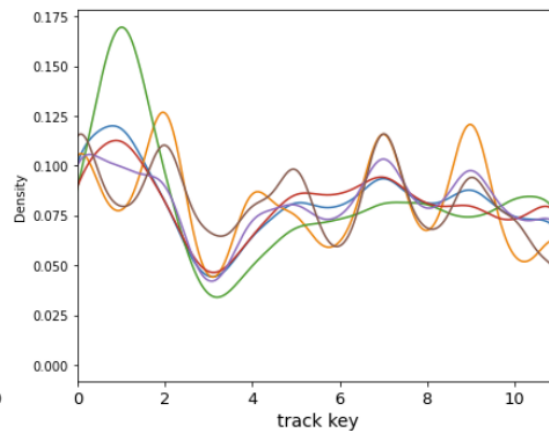
Min dolzina: 52827ms (52s) (0 min)



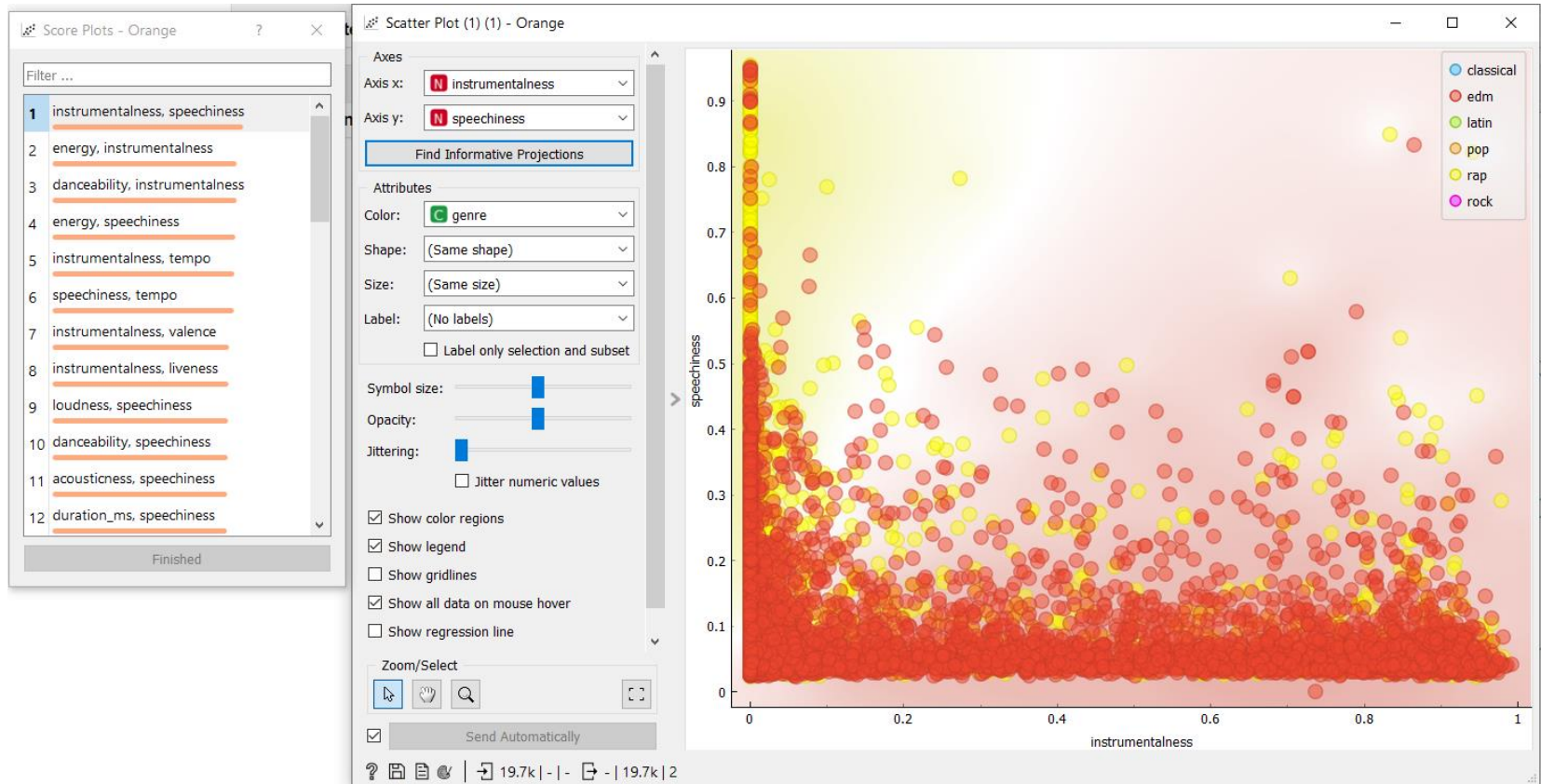
duration

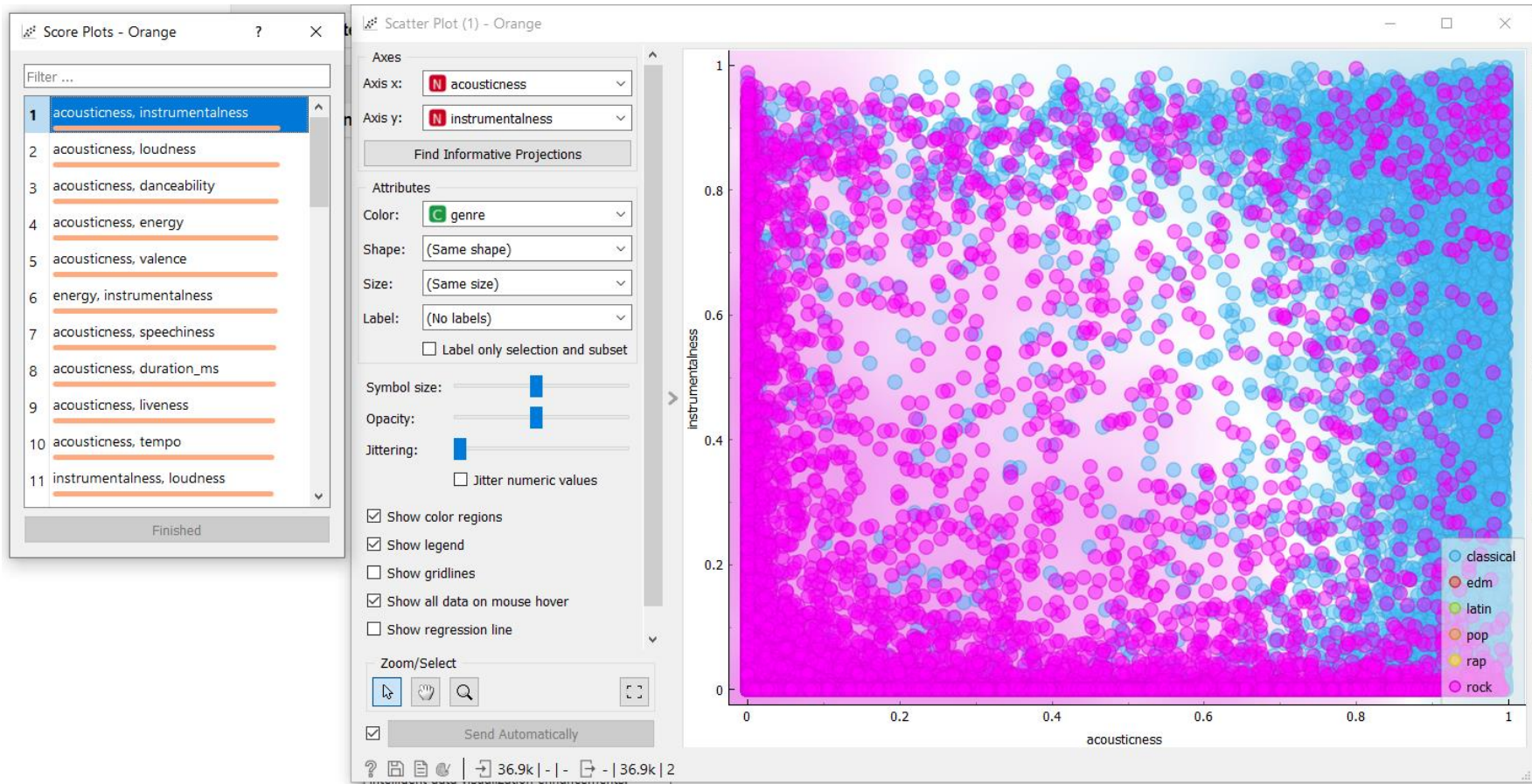


key



Povprečna dolžina pesmi  
200s = 3.3 min



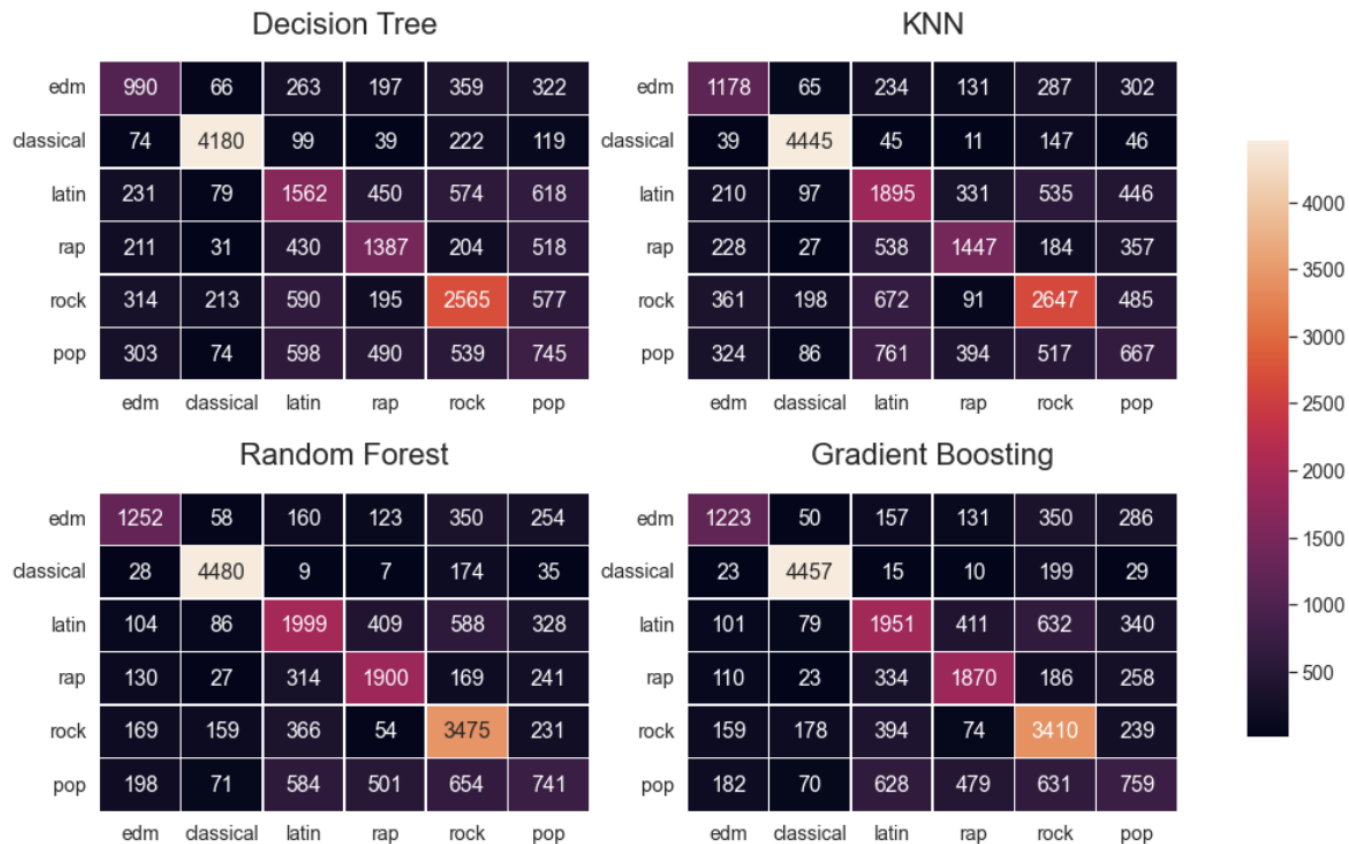


# Klasifikacija

```
Classifiers accuracy
Decision Tree: 0.56 (56%)
KNN: 0.59 (59%)
Random Forest: 0.67 (67%)
Gradient Boosting: 0.66 (66%)
```

Random Forest

	precision	recall	f1-score	support
classical	0.91	0.95	0.93	4704
edm	0.68	0.58	0.62	2205
latin	0.59	0.57	0.58	3479
pop	0.41	0.27	0.33	2782
rap	0.63	0.69	0.66	2775
rock	0.65	0.78	0.71	4483
accuracy			0.68	20428
macro avg	0.64	0.64	0.64	20428
weighted avg	0.67	0.68	0.67	20428





# 3. Sistem napovedovanja

## 3.1 podobnost vektorjev

1. Izracun vektorja - povprecje (mean) ALI sestevek (sum) vseh vrstic
2. Dodamo vektor deleza zanra (klasifikacijski model)
3. Primerjava dveh vektorjev s funkcijo cosine\_similarity (primerja smer 2 vektorjev)

```
# PLAYLISTS
# 0,1,2,3 - classical
# 4,5,6,7 - rap
# 8,9 - edm
# 10 - techno
# 11 - House
# 12,13,14,15 - rock
# 16,17,18,19 - latin
# 20,21,22,23 - pop
```

### Vektor lastnosti

playlist\_vector1\_sum

danceability	36.977300
energy	15.234030
loudness	69.692247
speechiness	6.565600
acousticness	135.374000
instrumentalness	108.083291
liveness	17.948000
valence	22.355400
tempo	71.024276
duration_ms	40.823697
dtype: float64]	

### Vektor deleza zanrov

Delez zanrov v playlistu 0

pop: 0.00  
rock: 0.01  
rap: 0.00  
edm: 0.00  
latin: 0.00  
classical: 0.99

**Klasika**

Delez zanrov v playlistu 5

pop: 0.13  
rock: 0.00  
rap: 0.70  
edm: 0.00  
latin: 0.17  
classical: 0.00

**Rap**

Podobnost playlistov:

Klasika - klasika: 1.00

Klasika - edm: 0.40

Rap - rap: 0.99

Rap - pop: 0.91

Rap - latin: 0.86

Rap - rock: 0.83

Rap - edm: 0.80

Rap - klasika: 0.39

edm - pop: 0.91

edm - rock: 0.88

edm - latin: 0.78

## 3.2 Napoved pesmi uporabnikom (KNN)

1. Naključno izberemo  $n$  - pesmi, te pesmi bomo napovedovali uporabnikom
2. Za vsako pesem izračunamo 200 najbliznjih sosedov vseh uporabnikov
3. Izračunamo pri katerih uporabnikih se pojavi največ sosed (delez)

```
# PLAYLISTS
# 0,1,2,3 - classical
# 4,5,6,7 - rap
# 8,9 - edm
# 10 - techno
# 11 - House
# 12,13,14,15 - rock
# 16,17,18,19 - latin
# 20,21,22,23 - pop
```

### Klasika

Pesem: The Blue Danube, Op. 314  
10 najbolj podobnih pesmi:

PL: 3 Song name: Die Fledermaus: Overture  
PL: 1 Song name: Die Zauberflöte (The Magic Flute), K. 620: Overture  
PL: 0 Song name: Il barbiere di Siviglia: Overture (Sinfonia)  
PL: 0 Song name: Merry Christmas Mr. Lawrence (Arr. Metcalfe)  
PL: 2 Song name: Symphony No.5 In C Minor, Op.67: 1. Allegro con brio  
PL: 0 Song name: Symphony No. 5 in C Minor, Op. 67: I. Allegro con brio  
PL: 1 Song name: Symphony No. 40 in G Minor, K. 550: I. Allegro molto  
PL: 0 Song name: Symphony No. 25 in G minor, K.183: 1. Allegro con brio  
PL: 1 Song name: Symphony No. 5 in C Minor, Op. 67 "Fate": Allegro con brio

### Rap

Pesem: Overseas  
10 najbolj podobnih pesmi:

PL: 6 Song name: Knock Knock  
PL: 4 Song name: Lifetime  
PL: 6 Song name: Reggae & Calypso (Russ Millions x Buni x YV)  
PL: 4 Song name: Mountains (with Fivio Foreign & Young M.A)  
PL: 6 Song name: John Wick  
PL: 5 Song name: 2055  
PL: 5 Song name: Scrub  
PL: 6 Song name: If I Die Young  
PL: 6 Song name: Slut

### Klasika

Playlist: 1 Delez najbliznjih sosedov: 64%  
Playlist: 2 Delez najbliznjih sosedov: 61%  
Playlist: 0 Delez najbliznjih sosedov: 48%

### Rap

Playlist: 7 Delez najbliznjih sosedov: 67%  
Playlist: 6 Delez najbliznjih sosedov: 64%  
Playlist: 4 Delez najbliznjih sosedov: 58%

## 3.3 Sistem napovedovanja (KNN)

1. Iskanje podobnih playlistov (KNN mean) - izberemo 3 najbolj podobne
2. Izbira bliznjih/podobnih pesmi (KNN) - izberemo 2 pesmi iz vsakega (podobnega) playlista
3. Predlagamo 6 najbolj podobnih pesmi (3 PL \* 2 pesmi)

<b>1. Podobni playlisti (uporabniki)</b>	Playlist 0  Najbolj podobni playlist: Playlist: 2 (razdalja 0.17) Playlist: 1 (razdalja 0.21) Playlist: 3 (razdalja 0.24)
<b>2. Razdalje + vrstice podobnih pesmi</b>	Razdalja, vrstica playlista (pesem) Razdalja: 0.137    (vrstica: 13) Razdalja: 0.141    (vrstica: 9) Razdalja: 0.149    (vrstica: 29)
<b>3. Ime pesmi data_set -&gt; vrstica</b>	Ave Maria, S. 558 (after Schubert, D. 839) Nocturne No.1 In B Flat Minor, Op.9 No.1 Pavane, Op.50

# Rezultati

# PLAYLISTS  
# 0,1,2,3 - classical  
# 4,5,6,7 - rap  
# 8,9 - edm  
# 10 - techno  
# 11 - House  
# 12,13,14,15 - rock  
# 16,17,18,19 - latin  
# 20,21,22,23 - pop

Recommended songs for playlist: 1

Name: River Flows in You - Orchestra Version || Artist: Yiruma  
Name: Symphony No.5 In C Minor, Op.67: 1. Allegro con brio || Artist: Ludwig van Beethoven  
Name: The Promise || Artist: Michael Nyman  
Name: Symphony No. 5 in C Minor, Op. 67: I. Allegro con brio || Artist: Ludwig van Beethoven  
Name: Die Walkure, Act III: Ride of the Valkyries || Artist: Richard Wagner  
Name: Finlandia, Op. 26 || Artist: Jean Sibelius

Recommended songs for playlist: 18

Name: Caraluna || Artist: Bacilos  
Name: Mi primer millon || Artist: Bacilos  
Name: Ley Seca || Artist: Jhay Cortez  
Name: Me Enamora || Artist: Mau y Ricky  
Name: MAMIII || Artist: Becky G  
Name: X ÚLTIMA VEZ || Artist: Daddy Yankee

Recommended songs for playlist: 4

Name: Supreme || Artist: Rippa da Kid  
Name: Authentic || Artist: Fayn  
Name: Over The Top (feat. Drake) || Artist: Smiley  
Name: Solid (feat. Drake) || Artist: Young Stoner Life  
Name: PENDEJO || Artist: Enrique Iglesias  
Name: GIRL LIKE ME || Artist: Black Eyed Peas

END