

Рационализируя иррациональное

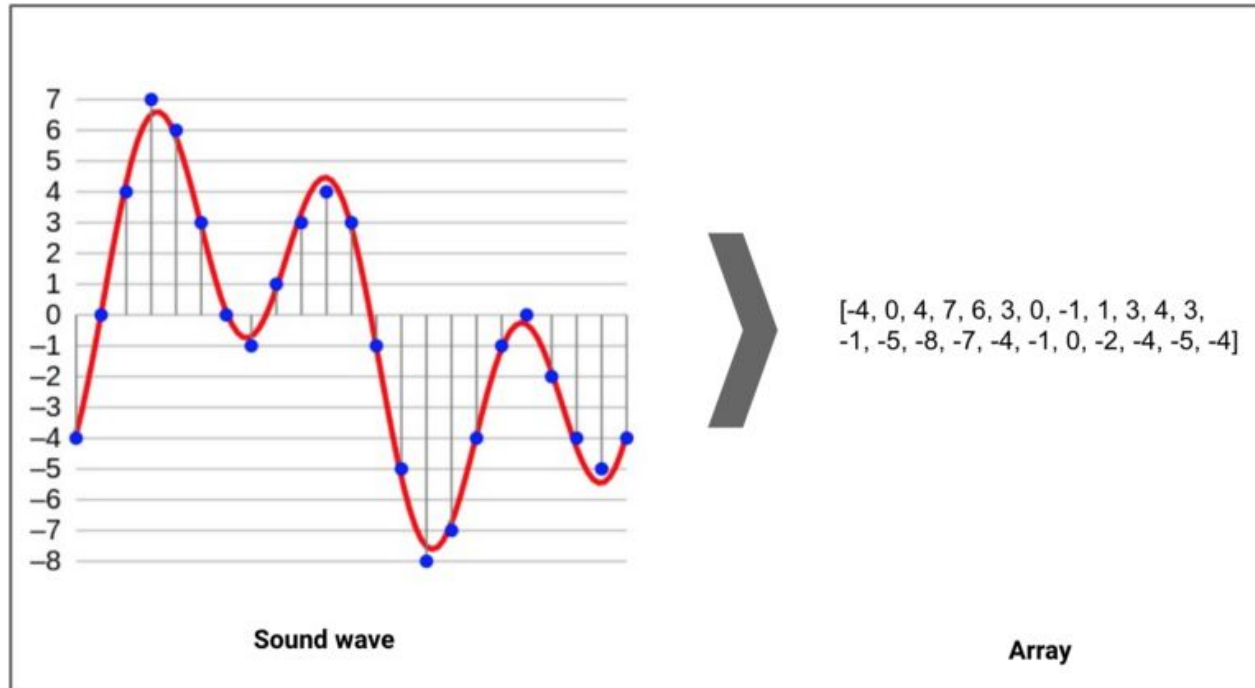
Дарья Никанорова
Евгения Стребулаева
Анна Рыбина
Екатерина Чернявская
Максим Сердаков

Methods

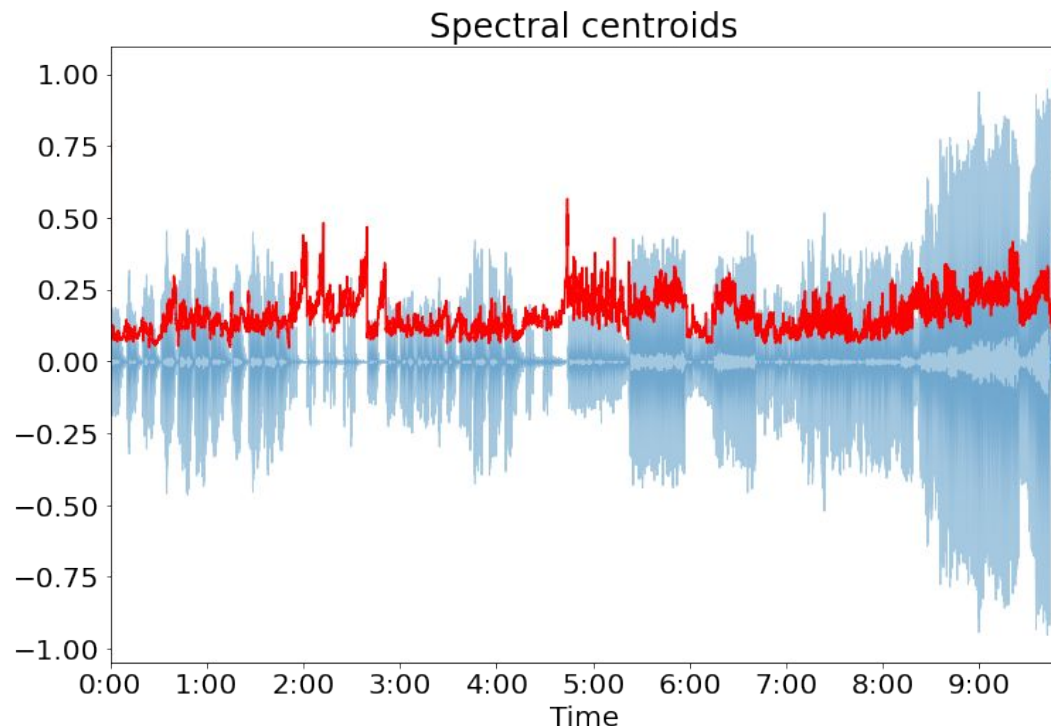
Feature extraction	librosa
Clusterization	PCA K-means
Classification	Random Forest KNN: k-nearest neighbours Keras sequential model
Multilabel classification	Keras



* Extra songs for better classification were obtained from
[GTZAN Dataset - Music Genre Classification](#)

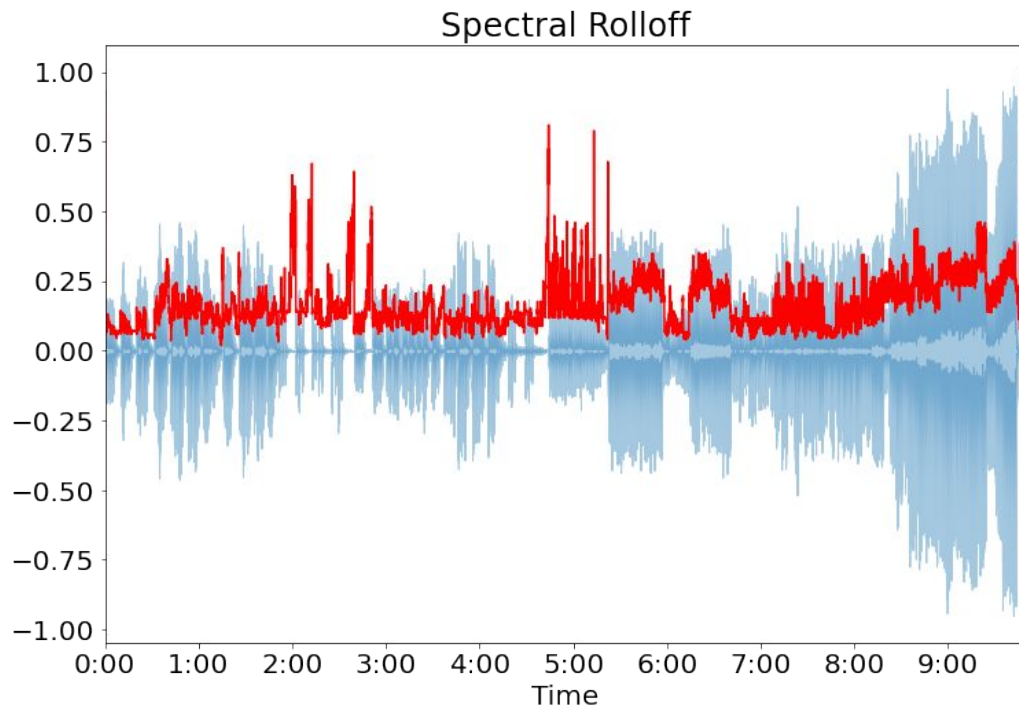


Непрерывный сигнал - в спектр дискретных значений



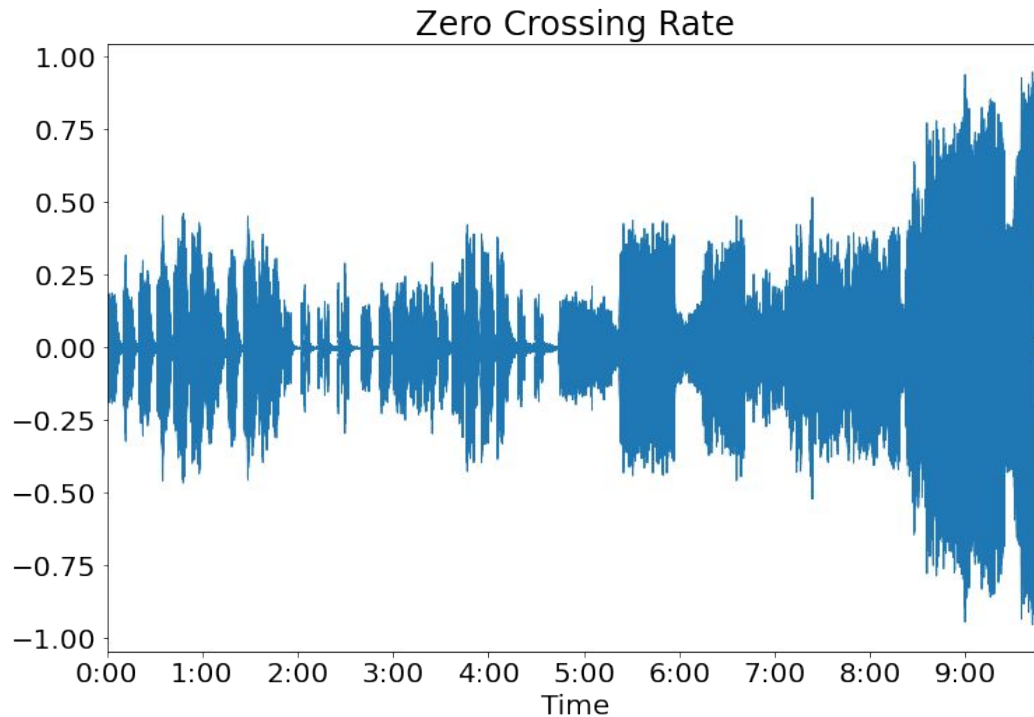
Спектральный центроид - указывает, где находится центр масс спектра. В восприятии он имеет прочную связь с впечатлением от яркости звука.

Спектральный центроид для каждого фрейма сигнала в Light_Of_The_Seven-Ramin_Djawadi.mp3



Это мера формы сигнала, представляющая собой частоту, в которой высокие частоты снижаются до 0. Или относительная частота, в пределах которой которой сосредоточена определенная часть всей энергии спектра (85% по умолчанию).

Частота спада для каждого фрейма в сигнале для Light_Of_The_Seven-Ramin_Djawadi.mp3



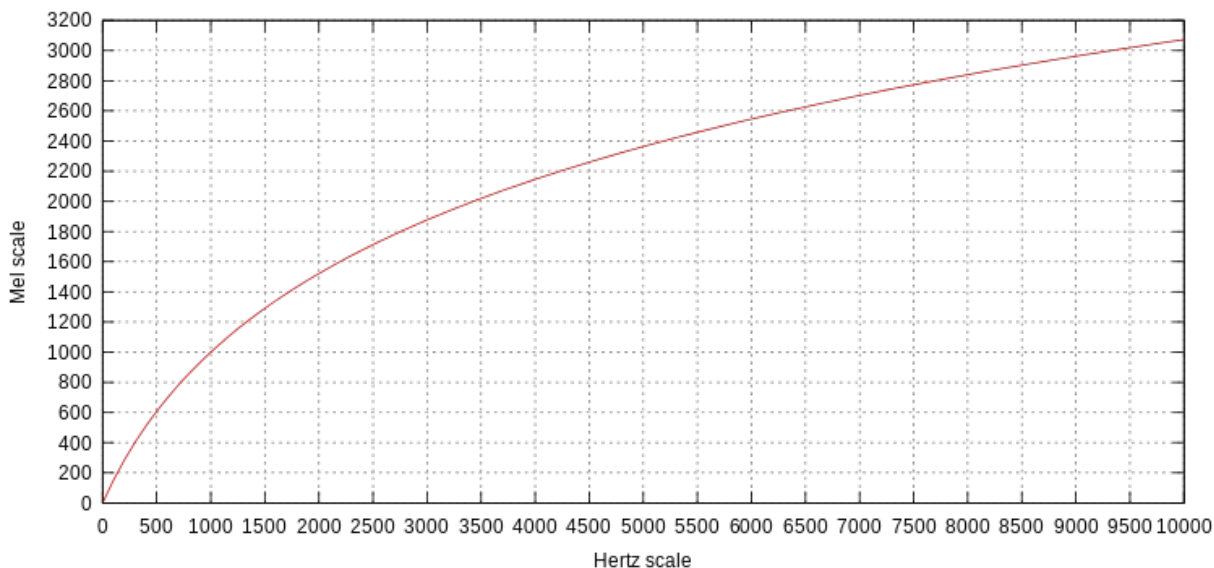
Вычисление гладкости сигнала - сигнал голоса колеблется довольно медленно, в роке и металле - высокие значения из-за высоких ударных звуков

Визуализация процессов пересечения нуля в Light_Of_The_Seven-Ramin_Djawadi.mp3

MFCC — Mel-Frequency Cepstral Coefficients



Мел - единица высоты звука, основанная на восприятии звука человеком.
Высота воспринимаемого звука не совсем линейно зависит от частоты:

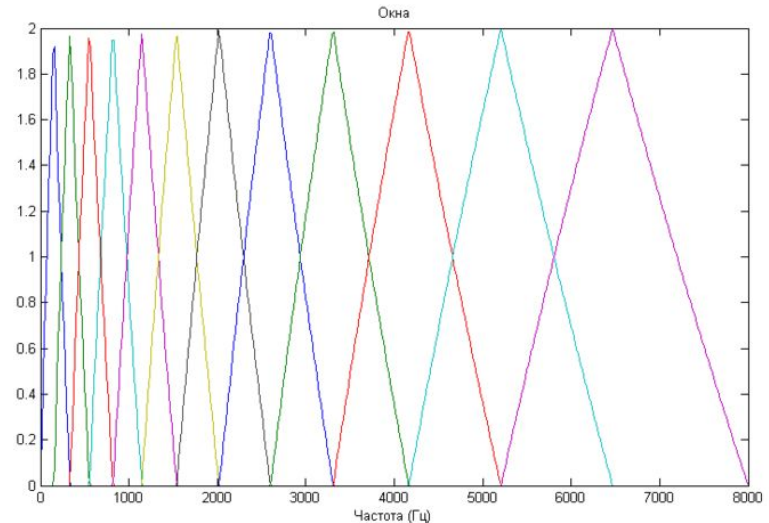
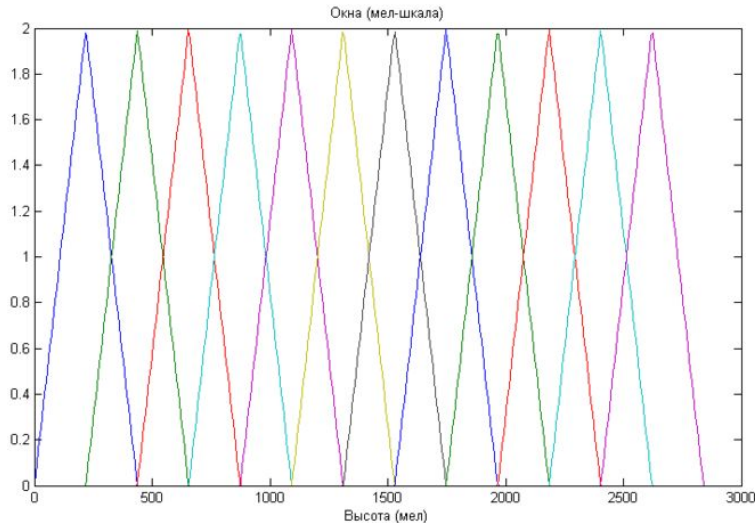


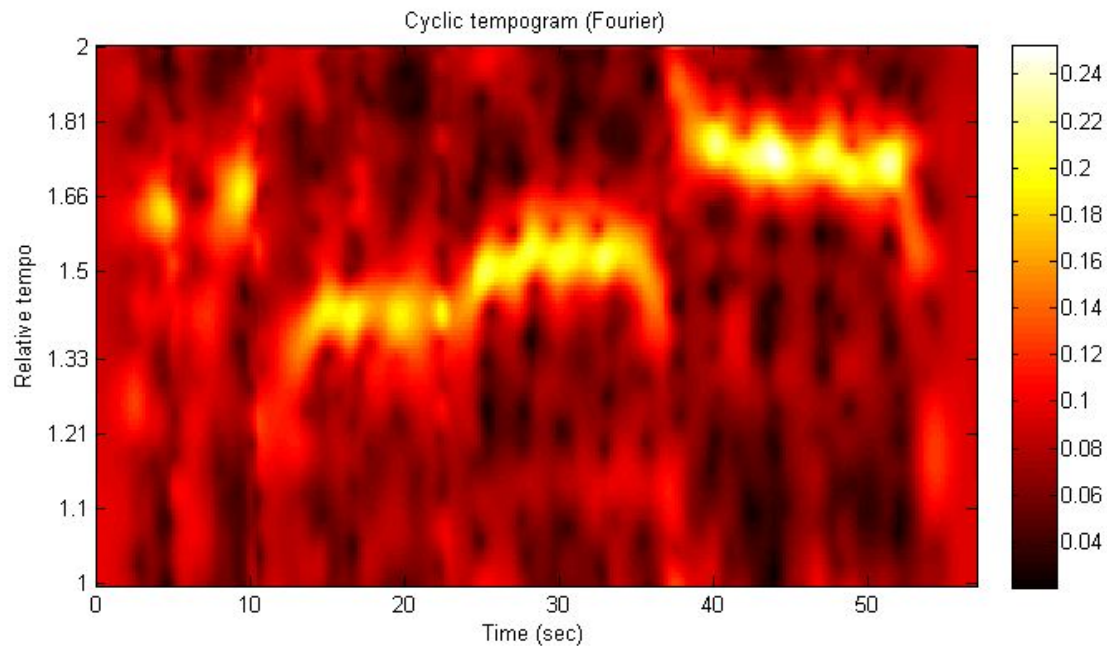
$$m = 2595 \log_{10} \left(1 + \frac{f}{700} \right) = 1127 \ln \left(1 + \frac{f}{700} \right)$$

MFCC — Mel-Frequency Cepstral Coefficients



Кепстр - это результат дискретного косинусного преобразования от логарифма амплитудного спектра сигнала. Звучит страшно, но, как итог, можно сказать, что мел-кепстральные коэффициенты - это спектральная огибающая с коэффициентами, расположенными друг от друга на расстоянии по шкале Мела, которые сосредоточены на важных для слуха частотах.

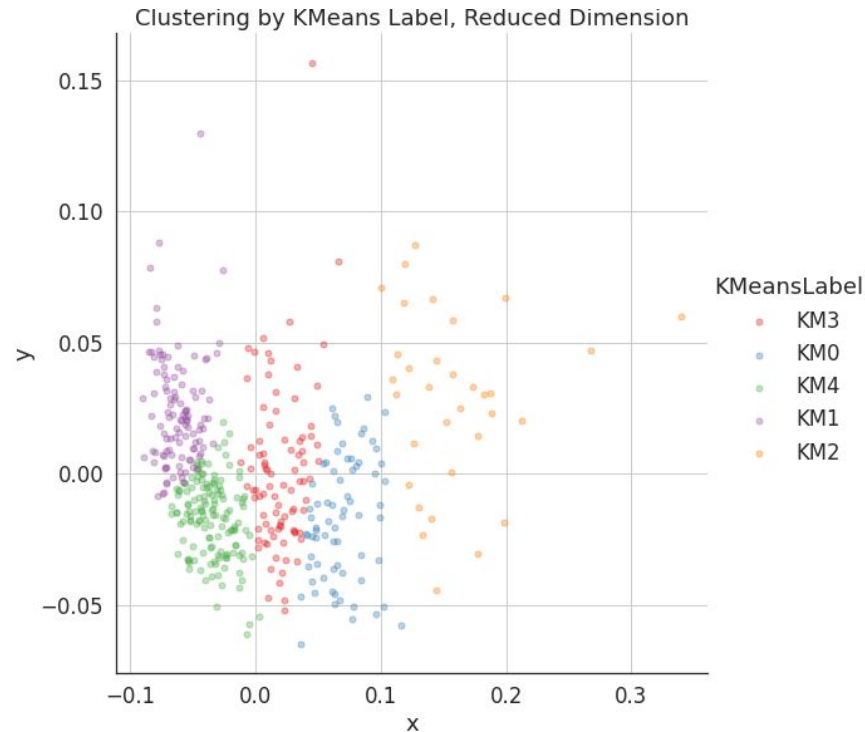
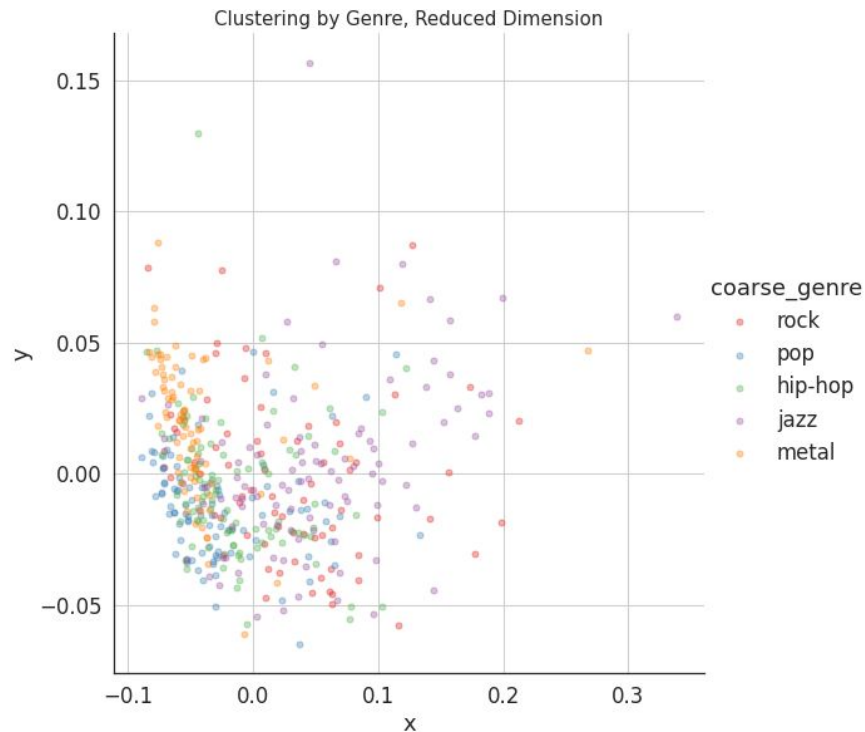




Оценка локальной
информации о
ритме и темпе

K-means clusterization

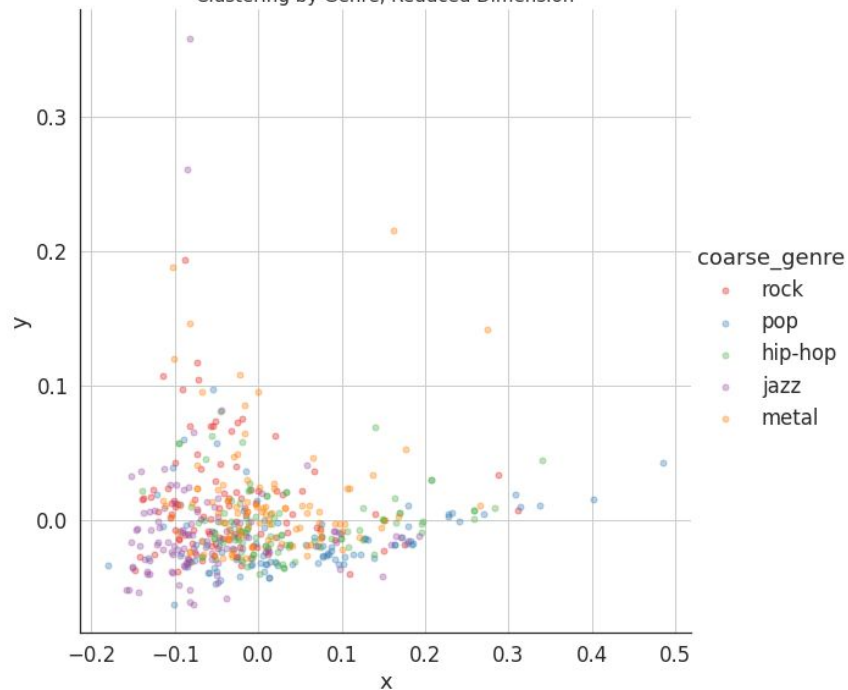
5 genres, partial feature list



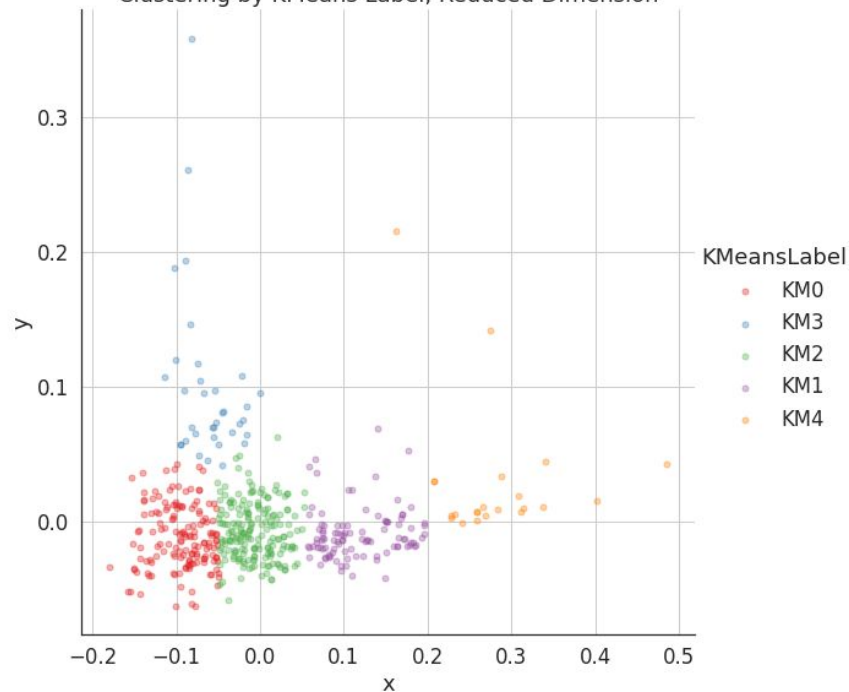
K-means clusterization

5 genres, full feature list

Clustering by Genre, Reduced Dimension

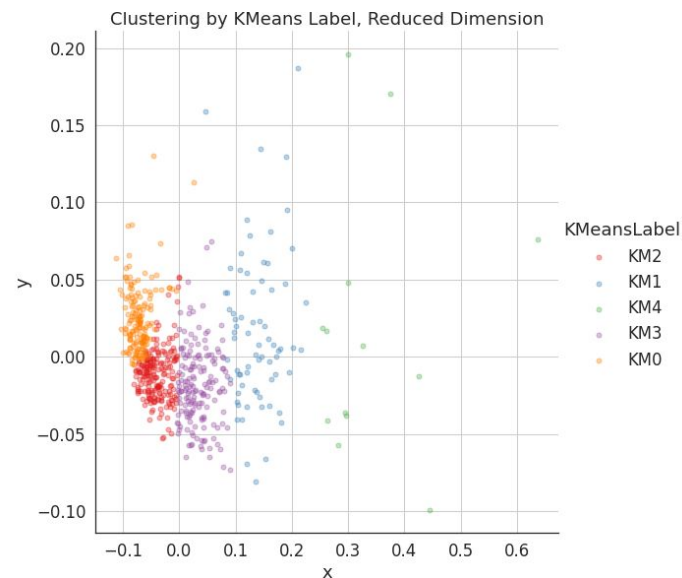
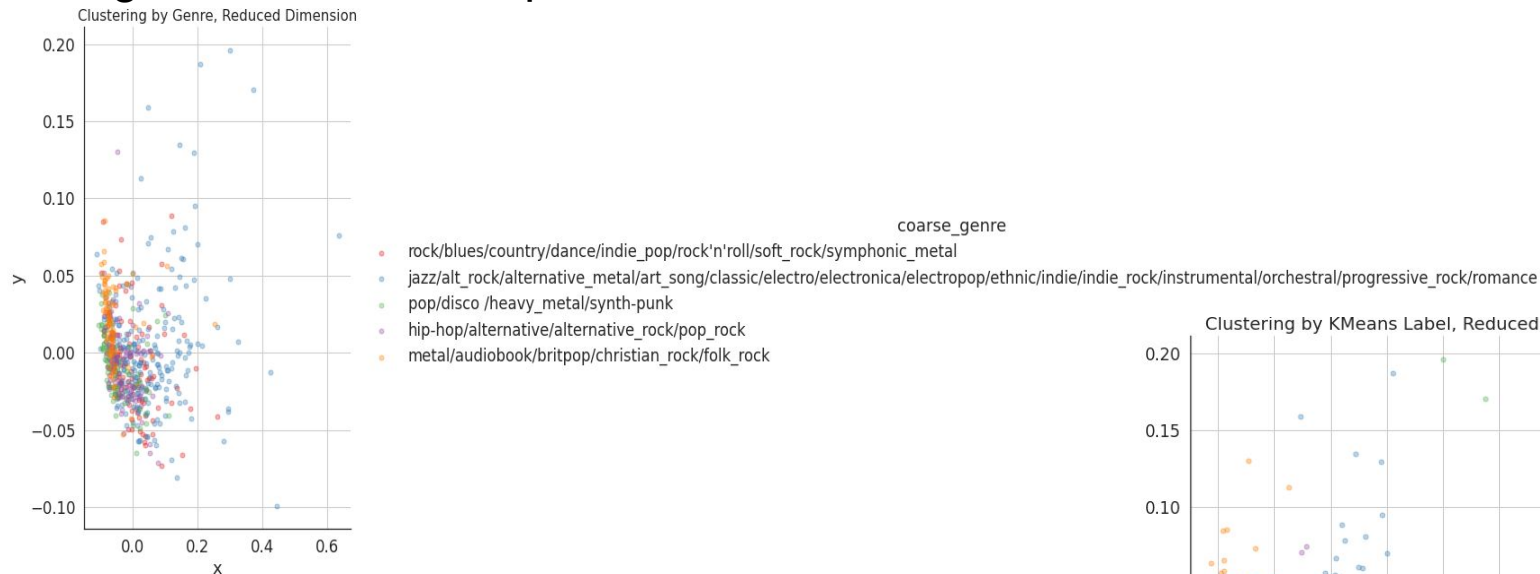


Clustering by KMeans Label, Reduced Dimension



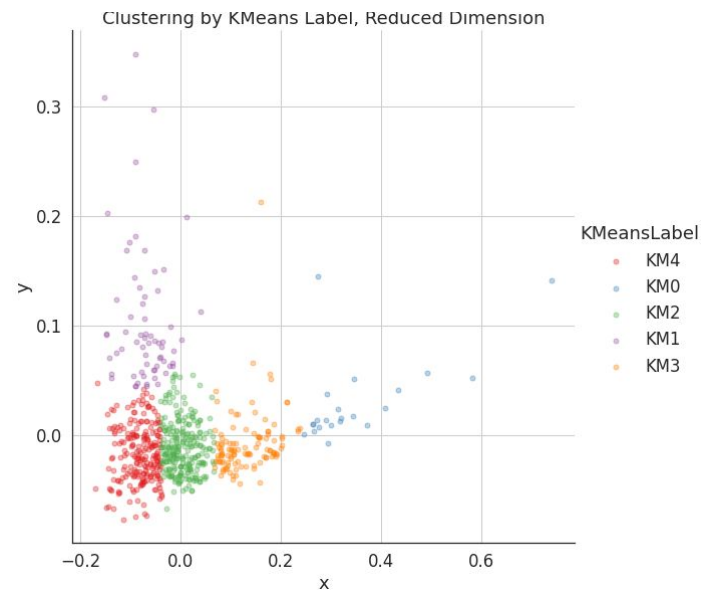
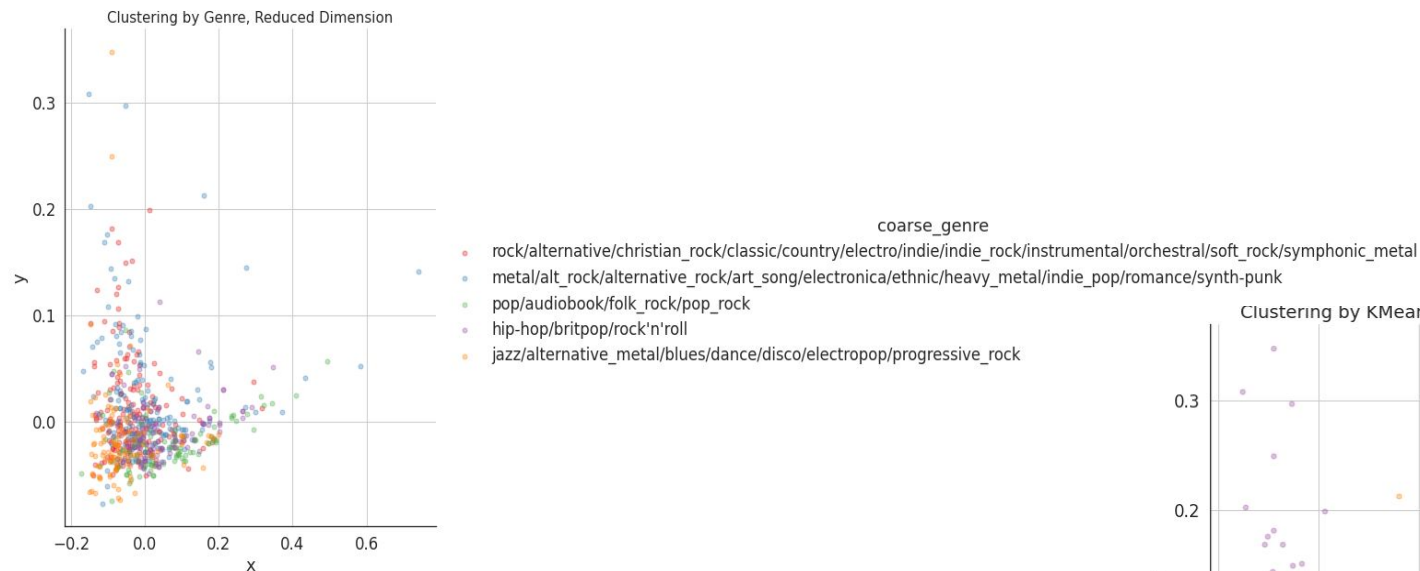
K-means clusterization

all genres united in 5, partial feature list

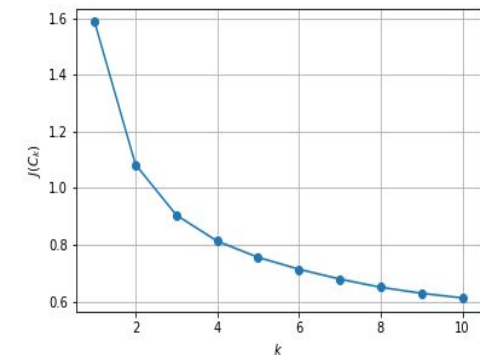


K-means clusterization

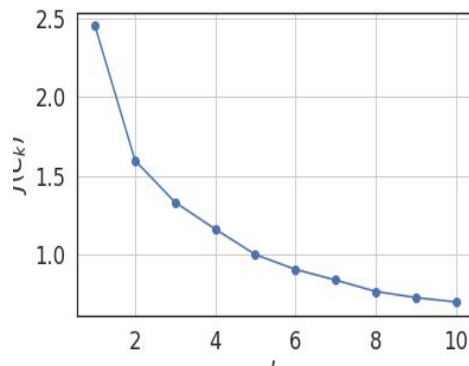
all genres united in 5, full feature list



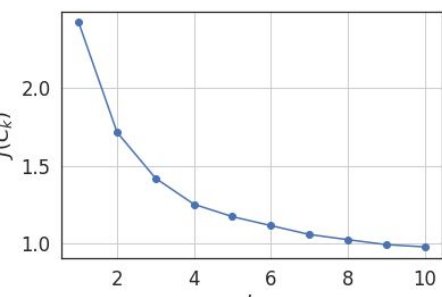
K-means: clusters & metrics



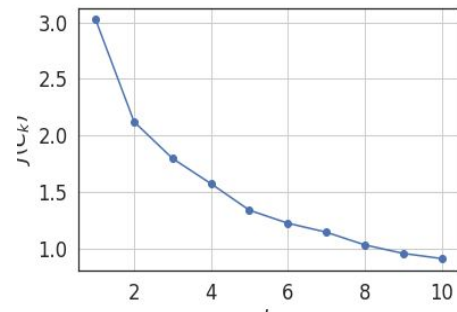
5 genres, partial



5 genres, full



all genres, partial

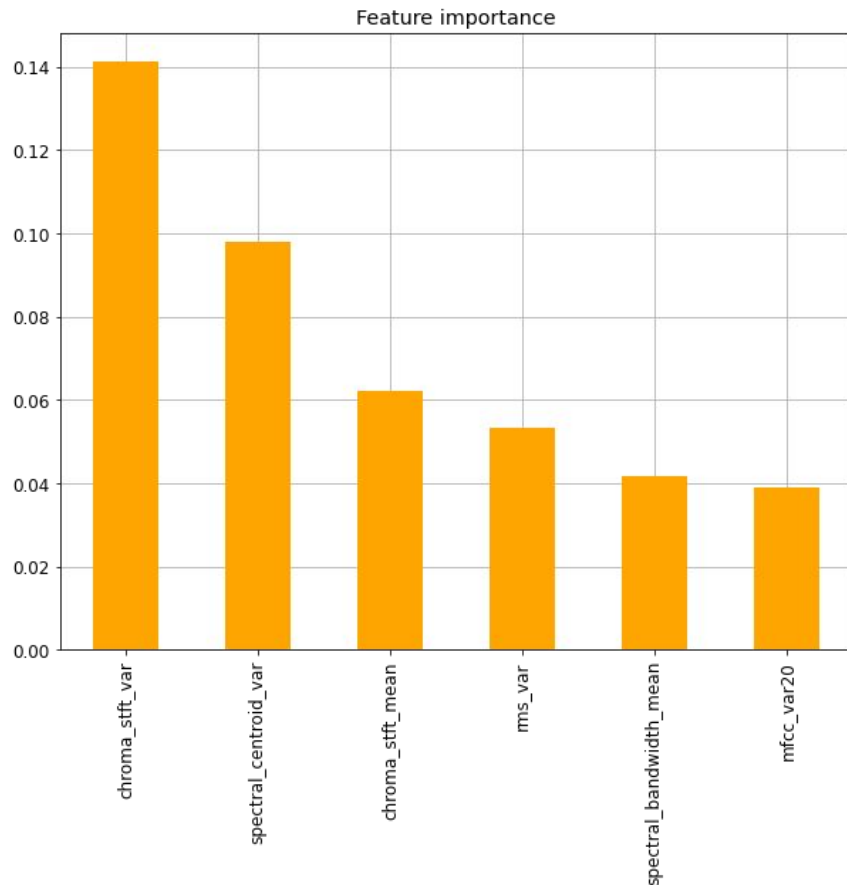


all genres, full

Instance	adjusted rand	adjusted mutual info	v measure	silhouette
5 genres, partial	0.0923	0.1209	0.1312	0.2982
5 genres, full	0.0753	0.0856	0.0955	0.4485
all genres, partial	0.0691	0.12658	0.1347	0.3096
all genres, full	0.0607	0.0829	0.0906	0.4190

- Adjusted Rand Score: Bounded $[-1,1]$ where 1 represents perfect label, and a score near or below 0 have independent, or "bad" labelings.
- Adjusted Mutual Info Score: Bounded $[0,1]$ where 1 represents perfect label, and a score near 0 have independent, or "bad" labelings
- V Measure Score: Bounded $[0,1]$ scores: 0.0 is as bad as it can be, 1.0 is a perfect score.
- Silhouette Score: The score is bounded between -1 for incorrect clustering and +1 for highly dense clustering. Scores around zero indicate overlapping clusters.

Random Forest classifier



Parameters

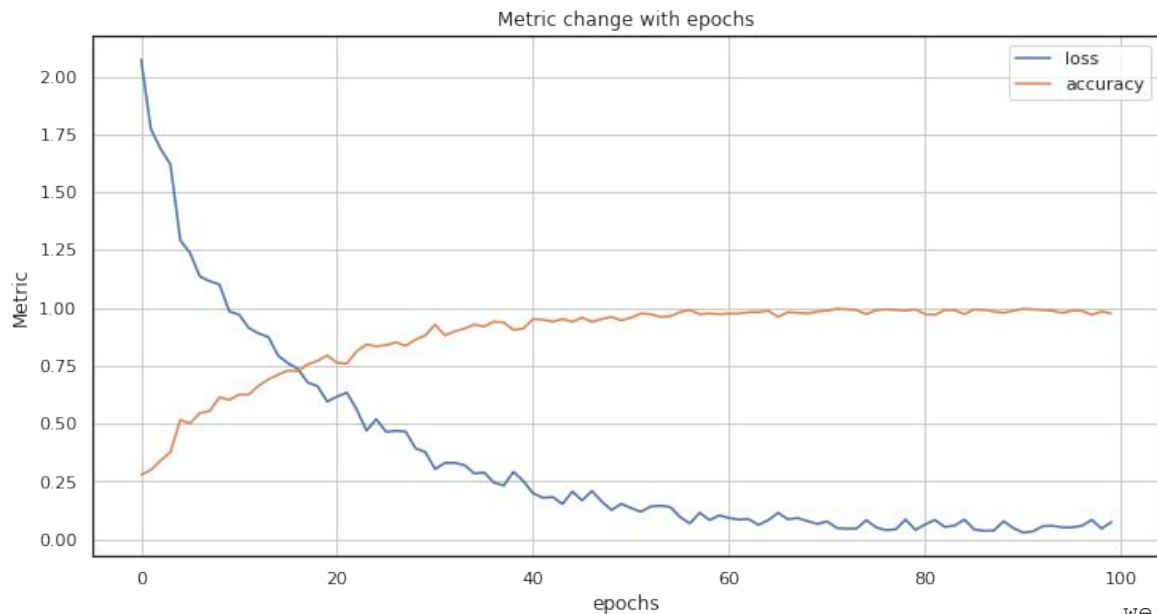
```
max_depth: 7,  
min_samples_leaf: 2,  
min_samples_split: 5,  
n_estimators: 25
```

Metrics

	precision	recall	f1-score	support
0	0.65	0.61	0.62	33
1	0.94	0.88	0.91	33
2	0.77	0.71	0.74	34
3	0.74	0.79	0.76	33
4	0.45	0.52	0.48	33
accuracy			0.70	166
macro avg	0.71	0.70	0.70	166
weighted avg	0.71	0.70	0.70	166

Keras sequential model

epochs=100, no cross-validation, single run



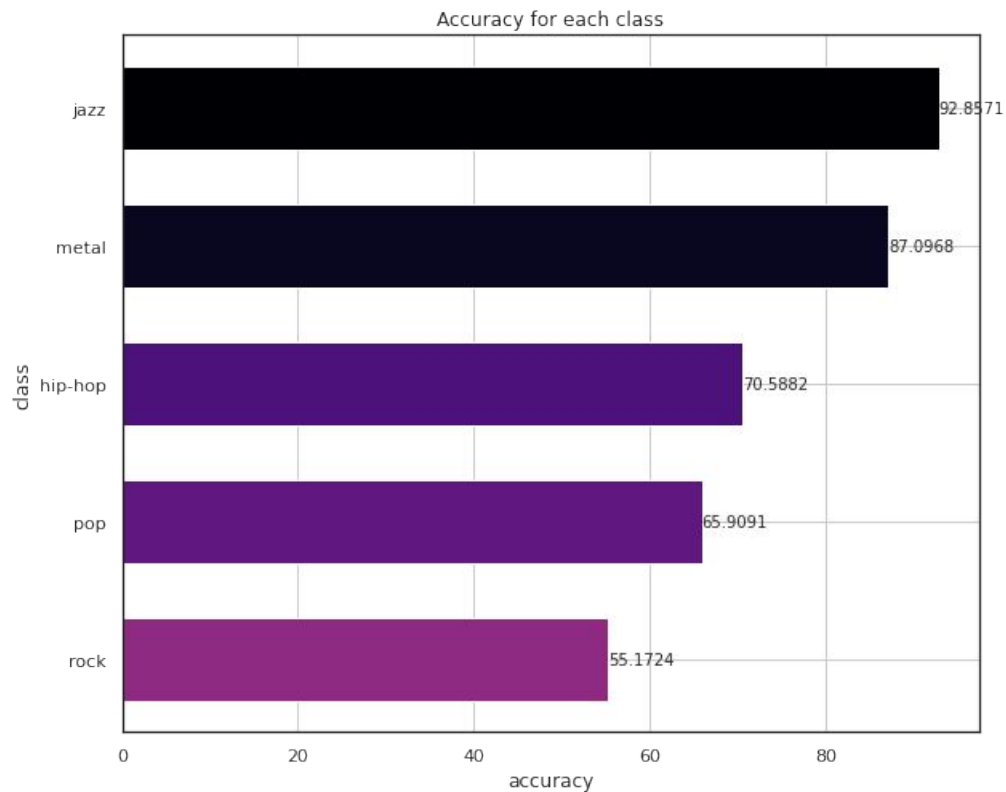
Parameters

```
optimizer='rmsprop'  
loss='categorical_crossentropy'
```

Metrics (classification report for 1 run)

	precision	recall	f1-score
0	0.71	0.67	0.69
1	0.93	0.87	0.90
2	0.87	0.82	0.84
3	0.66	0.85	0.74
4	0.55	0.48	0.52
accuracy			0.73
macro avg	0.74	0.74	0.74
weighted avg	0.74	0.73	0.73

Keras sequential model



Parameters

```
optimizer='rmsprop'  
loss='categorical_crossentropy'
```

Metrics

	precision	recall	f1-score
0	0.71	0.67	0.69
1	0.93	0.87	0.90
2	0.87	0.82	0.84
3	0.66	0.85	0.74
4	0.55	0.48	0.52
accuracy			0.73
macro avg	0.74	0.74	0.74
weighted avg	0.74	0.73	0.73

Keras sequential model

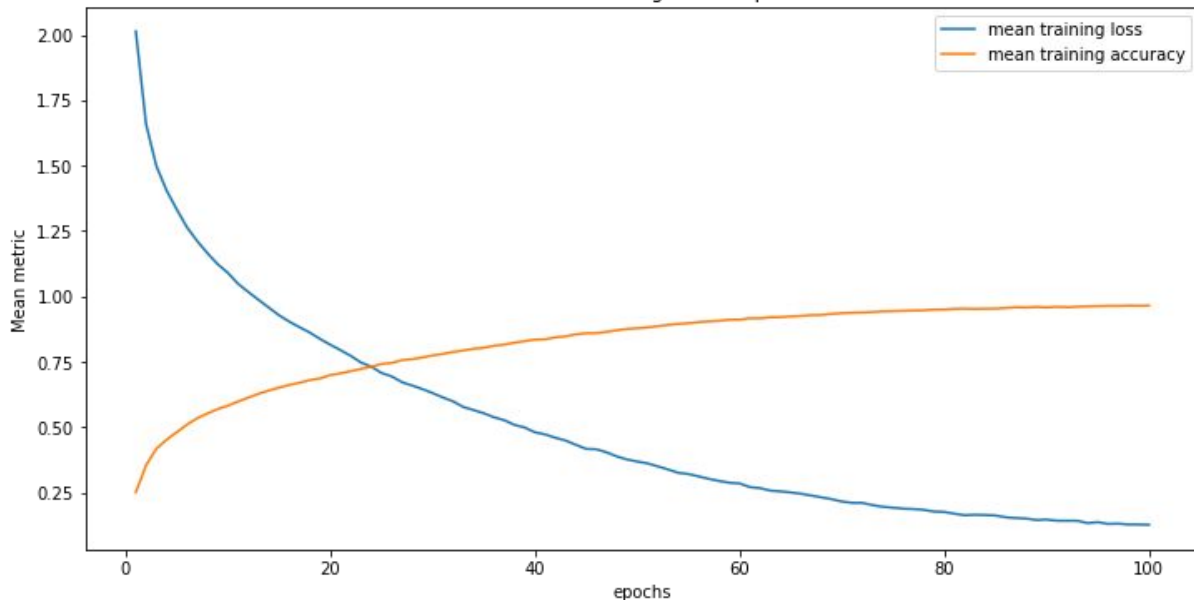
epochs=100, cv=5, repeats=30

Parameters

```
optimizer='rmsprop'  
loss='categorical_crossentropy'
```

Mean test accuracy: 0.769
The 0.95% conf. interval: [0.745, 0.793]

Mean metric changes with epochs



KNN: k-nearest neighbours



Metrics

	precision	recall	f1-score	support
hip-hop	0.69	0.73	0.71	30
jazz	0.88	0.73	0.80	30
metal	0.81	0.84	0.83	31
pop	0.61	0.73	0.67	30
rock	0.58	0.50	0.54	30
accuracy			0.71	151
macro avg	0.71	0.71	0.71	151
weighted avg	0.71	0.71	0.71	151

no prior PCA

- with/without prior PCA
- cv=5, GridSearchCV
- best parameters:

knn__algorithm: auto

knn__n_neighbors: 8

knn__weights: distance

	precision	recall	f1-score	support
hip-hop	0.77	0.67	0.71	30
jazz	0.73	0.63	0.68	30
metal	0.76	0.84	0.80	31
pop	0.65	0.87	0.74	30
rock	0.60	0.50	0.55	30
accuracy			0.70	151
macro avg	0.70	0.70	0.70	151
weighted avg	0.70	0.70	0.70	151

after PCA
(55 -> 29 features)

Multilabel classification

Table 1

X	y
X_1	t_1
X_2	t_2
X_3	t_1
X_4	t_2
X_5	t_1

Binary Classification

Table 2

X	y
X_1	t_2
X_2	t_3
X_3	t_4
X_4	t_1
X_5	t_3

Multi-class Classification

Table 3

X	y
X_1	$[t_2, t_5]$
X_2	$[t_1, t_2, t_3, t_4]$
X_3	$[t_3]$
X_3	$[t_2, t_4]$
X_3	$[t_1, t_3, t_4]$

Multi-label Classification

	alternative	electronica	hip-hop	metal	pop	rock
0	0	0	0	0	0	1
1	0	0	0	0	0	0
2	0	0	0	0	0	0
3	0	1	0	0	0	0
4	0	1	0	0	0	0

Predicted sub genres:

Train dataframe:

song	alternative	electronica	hip-hop	metal	pop	rock
7Horse_-_Answer_the_Bell.mp3	0	0	0	0	0	1
A_Simple_Life-Brian_Crain.mp3	0	0	0	0	0	0
Adagio_in_G_Minor-Albinoni.mp3	0	0	0	0	0	0
Age_Atomic-Solar_Bears.mp3	0	1	0	0	0	0
Alakazam_I-Justice.mp3	0	1	0	0	0	0

Test dataframe:

Beyond_The_Black-Burning_in_Flames.mp3	0	0	0	0	0	1
Beyond_The_Black-Fairytale_of_Doom.mp3	0	0	0	0	0	1
Beyond_The_Black-Hysteria.mp3	0	0	1	0	0	0
Big_Twice_-_Super_Mario_Bros.mp3	0	0	0	0	0	1
Billy_Joel_-_Honesty_47998187.mp3	0	0	1	0	1	0

Parameters:

BinaryRelevance:

```
classifier=SVC()
```

Keras: (better)

```
activation='sigmoid'
loss='binary_crossentropy'
optimizer=sgd
```

Hamming loss: 0.2778

(still not good)

Ссылки

1. <https://medium.com/nuances-of-programming/%D0%B0%D0%BD%D0%B0%D0%B%D0%B8%D0%B7-%D0%B0%D1%83%D0%B4%D0%B8%D0%BE%D0%B4%D0%B0%D0%BD%D0%BD%D1%8B%D1%85-%D1%81-%D0%BF%D0%BE%D0%BC%D0%BE%D1%89%D1%8C%D1%8E-%D0%B3%D0%BB%D1%83%D0%B1%D0%BE%D0%BA%D0%BE%D0%B3%D0%BE-%D0%BE%D0%B1%D1%83%D1%87%D0%B5%D0%BD%D0%B8%D1%8F-%D0%B8-python-%D1%87%D0%B0%D1%81%D1%82%D1%8C-1-2056fef8525e>
2. <https://habr.com/ru/post/140828/>
3. <https://fobosworld.ru/kepstralnye-koeffitsienty-kak-neobhodimaya-harakteristika-prot sessa-sozdaniya-sistemy-imitatsii-golosa-cheloveka-s-pomoshhyu-metodov-glubokogo-obucheniya-mezhdunarodnyj-studencheskij-nauchnyj-vestnik/>