

Projekt klasyfikacja gatunków muzyki:

Link do repozytorium na github:

https://github.com/kochamkotkiic/music_genre_classification.git

1 etap projektu:

Podział zadań:

Emilia Urbanek:

1. Znaleźć i pobrać dane

- Pobrać dataset **GTZAN** z TensorFlow Datasets- Napotkaliśmy problem z pobraniem ich z database TensorFlow - według różnych źródeł jest on już niedostępny, a wszelkie inne znalezione alternatywy są niedostępne w Polsce(jest to kanadyjska strona- i dodatkowo ma zawodne serwery). Link do tej strony-<http://opihi.cs.uvic.ca/sound/genres.tar.gz>.
Udało nam się znaleźć offlinową wersję tego samego zbioru danych do manualnego pobrania- z tego rozwiązania będziemy korzystać w projekcie.
- Uruchomić kod ładowania danych.

2. Część EDA: opis cech- wstępny dokument z analizą

- Wyjaśnić jakie cechy można wyciągnąć z audio.
- Pokazać wykres: MFCC, RMS Energy, PCA, t-SNE
- Opisać co przedstawiają różne widoki sygnału audio.
- Policzyć ile jest utworów w każdej klasie (rock, jazz, classical, metal...).
- Sprawdzić długość nagrań.
- Przeanalizować rozkłady cech:

Mikołaj Frąckowiak

1. Utworzenie Dockera

Julia Dąbrowska

Jak wczytać dane?

W PowerShell (Windows) (przykładowe ścieżki):

```
cd C:\Users\emilk\music_genre_classification\docker-ścieżka do folderu projektu
```

Uruchomienie dockera z montowanym folderem (ścieżka do folderu z danymi)

```
docker-compose run --rm -v "C:\Users\emilk\data\GTZAN:/app/data" ml-project bash
```

Uruchomienie skryptu:

Wewnątrz kontenera:

```
python scripts/prepare_datasets.py --data-dir "/app/data"
```

Wyniki (data/processed) zapisują się w folderze projektu.

Po wczytaniu danych należy uruchomić kod treningu:

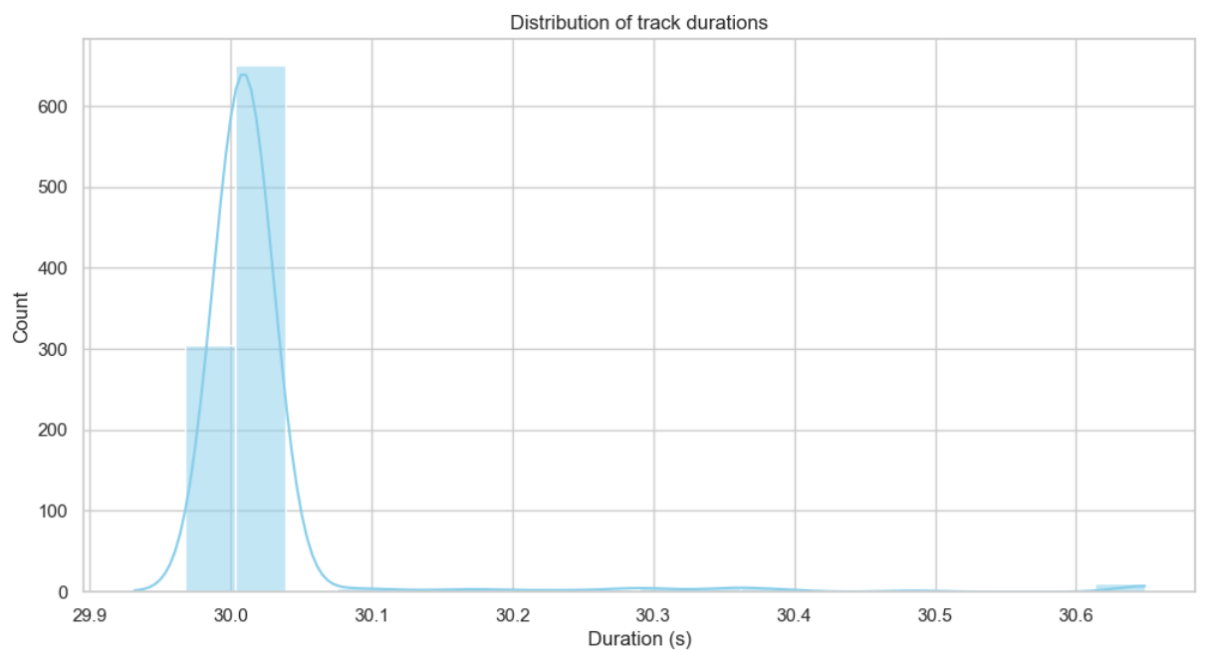
```
python scripts/train_knn.py
```

EDA:

Wczytywanie danych:

```
Genres found: ['blues', 'classical', 'country', 'disco', 'hiphop', 'jazz', 'metal', 'pop', 'reggae', 'rock']
audio genre filename
0 [0.0073242188, 0.016601562, 0.0076293945, -0.0... blues blues.00000.au
1 [0.0034179688, 0.0043029785, 0.001373291, 0.00... blues blues.00001.au
2 [0.019012451, 0.047698975, 0.029418945, -0.018... blues blues.00002.au
3 [-0.013000488, -0.03060913, -0.036071777, -0.0... blues blues.00003.au
4 [-0.0063171387, -0.009277344, -0.008331299, -0... blues blues.00004.au
Total samples: 1000
genre
blues 100
classical 100
country 100
disco 100
hiphop 100
jazz 100
metal 100
pop 100
reggae 100
rock 100
Name: count, dtype: int64
Class names: ['blues', 'classical', 'country', 'disco', 'hiphop', 'jazz', 'metal', 'pop', 'reggae', 'rock']
```

Długość nagrań:



```
Mean duration: 30.024074648526074
Min duration: 29.931972789115648
Max duration: 30.648888888888887
```

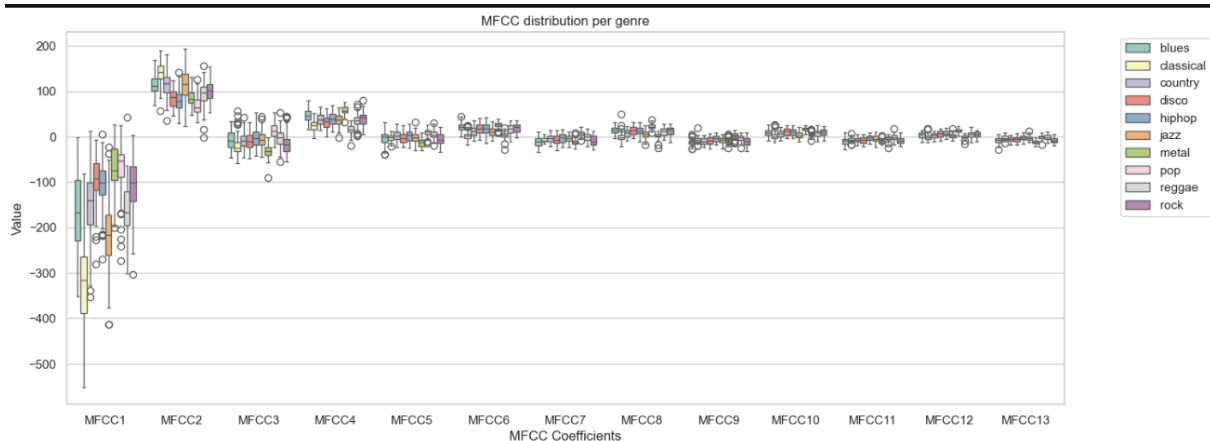
Feature extraction (MFCC, Chroma, RMS):

Features shape: (1000, 59)

1. **Liczba wierszy** w features_ext = liczba nagrań = 1000
2. **Liczba kolumn** = liczba cech wyciągniętych z jednego nagrania:
 - MFCC: 13 średnich + 13 odchyleń = 26
 - Chroma: 12 średnich + 12 odchyleń = 24
 - RMS: 1
 - Spectral Contrast: 7 średnich
 - Zero Crossing Rate: 1

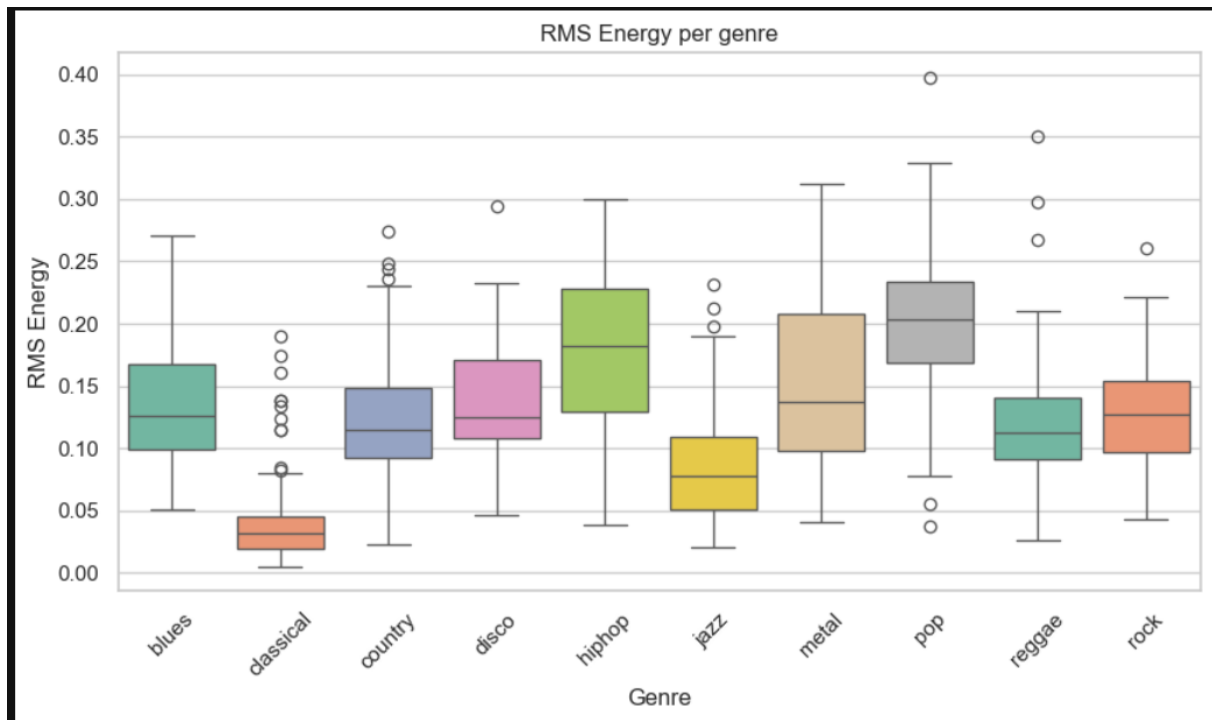
Łącznie: $26 + 24 + 1 + 7 + 1 = 59$ kolumn

MFCC Boxplot per genre:



Rozkłady MFCC ukazują, że pierwszy współczynnik (MFCC1-miernik zbliżony do „ogólnej energii / jasności” sygnału) odpowiada w dużej mierze za ogólną energię nagrań, podczas gdy MFCC2-13 przenoszą informacje o barwie dźwięku. Utwory klasyczne wyróżniają się niskimi wartościami MFCC1, co sugeruje, że model klasyfikacyjny może je odróżnić od pozostałych gatunków z wysoką skutecznością.

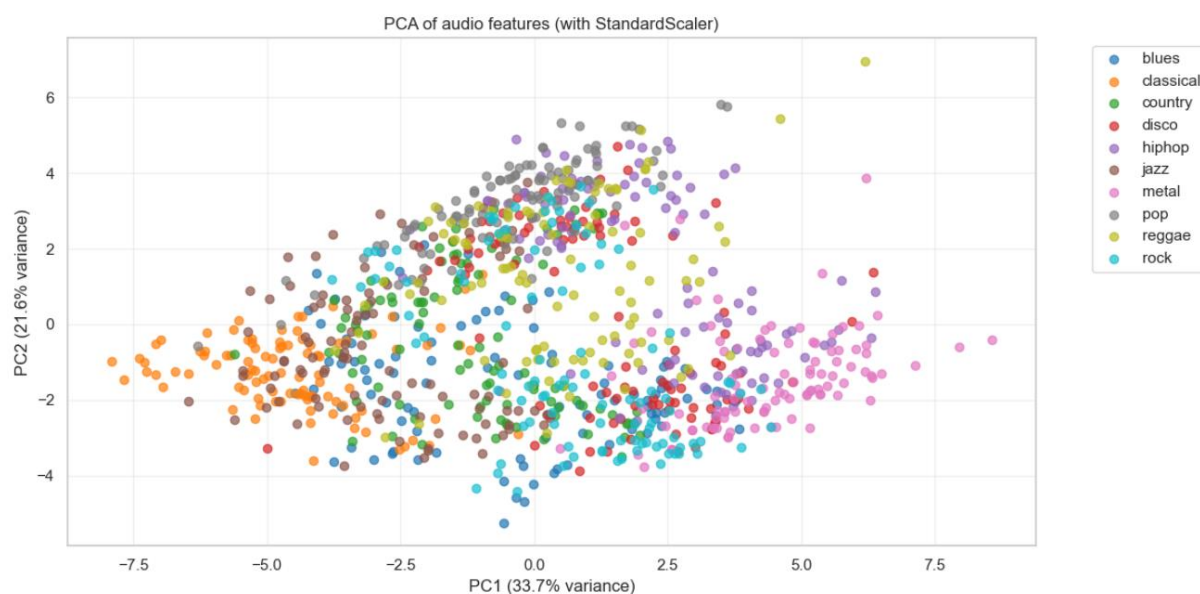
RMS Energy per genre:



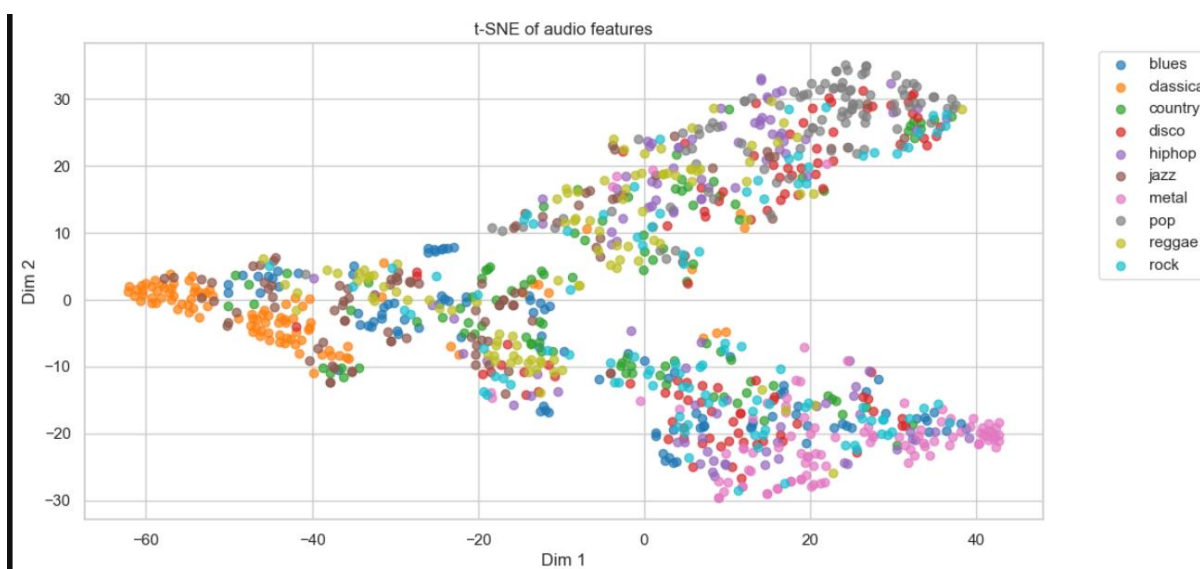
- **Mediana** w środku pudełka – typowy poziom głośności utworu w tym gatunku
- **Wysokość pudełka** – zmienność energii między utworami (czy gatunek ma podobną głośność w większości nagrań)
- **Długie wąsy/pudełko** – duża rozpiętość głośności (duża dynamika)
- **Krótkie wąsy/pudełko** – nagrania są do siebie podobne pod względem głośności
- **Kropki** – pojedyncze utwory nietypowo głośne/ciche

Analiza RMS Energy wykazała znaczące różnice między gatunkami. Muzyka klasyczna charakteryzuje się najniższą średnią energią oraz dużą dynamiką, co odzwierciedla naturalne ciche i głośne fragmenty utworów. Gatunki pop i hip-hop wykazują wysoką i stosunkowo stałą energię sygnału. Metal cechuje wysoka energia z większym rozrzutem niż pop.

PCA:



t-SNE:



Analiza redukcji wymiarów przy użyciu t-SNE pokazała naturalne skupiska gatunków. Klasyczne i jazzowe utwory tworzą wyraźnie odseparowane grupy, natomiast gatunki takie jak rock, country i disco nachodzą na siebie w przestrzeni cech, co wskazuje na potencjalne trudności w klasyfikacji. Metal i pop tworzą dość zwarte skupiska, z pewnymi nakładkami na inne gatunki. Wynika to z podobnego instrumentarium i struktury harmoniczej.

Podsumowanie analizy danych audio GTZAN :

W analizowanym zbiorze danych GTZAN znajduje się 1000 nagrań podzielonych równomiernie na 10 gatunków muzycznych (po 100 utworów na klasę). Długość nagrań jest spójna, średnio około 30 sekund (min. 29.9 s, max. 30.6 s), co eliminuje potrzebę dodatkowego przycinania lub dopełniania próbek.

Wnioski:

- Problem klasyfikacji jest rozwiązywalny, z przewidywaną dokładnością 60–80% dla prostych modeli opartych na MFCC.

- Najłatwiejszy do rozróżnienia gatunek: Classical.
- Najtrudniejsze do oddzielenia pary: Rock, Country i Blues, ze względu na podobne cechy widmowe i instrumentarium.