

# Klasyfikacja gatunków utworu na podstawie jego metadanych dostarczanych przez Spotify API

Adam Lamers

*Organization* Politechnika Wrocławska

Abstract

## Contents

### 1 Introduction

Klasyfikacja jest procesem predykcji dotyczącej klasy lub klas do jakich należy obiekt. Predykcja ta odbywa się na podstawie analizy danych obiektu, które mogą służyć do jego klasyfikacji. Posiadając zbiór takich obiektów jesteśmy w stanie na podstawie

zależności występujących w tym zbiorze (nazywanym zbiorem uczącym) jesteśmy w stanie przewidywać przynależności do danej klasy obiektów z poza zbioru uczącego. Wymagania dotyczące zbioru wachają się względem różnych modeli wykorzystywanych do klasyfikacji. W przypadku tego raportu, klasyfikacji podlegają utwory muzyczne na podstawie ich gatunku lub gatunków.

gray

---

## 2 Źródło danych

Wstępna analiza dostępnych danych i ich charakterystyki

gray

---

### 2.1 Spotify Web API

#### 2.1.1 Architektura Spotify Web API

Dane wykorzystane w tym raporcie pochodzą z Webowego API Spo-

tify przeznaczonego dla deweloperów third-party. Wykorzystane API umożliwia uzyskanie metadanych o utworze po podaniu ważnego tokena i identyfikatora utworu (`track_id`).

System uwierzytelniania i autoryzacji wykorzystany przez Spotify API to OAuth, w związku z tym aby uzyskać metadane utworów należy się wcześniej uwierzytelnić.

### **2.1.2 Implementacja wrappera Spotify Web API**

Aby zautomatyzować proces uwierzytelniania napisałem prosty wrapper, który umożliwia pobieranie danych z najważniejszych endpointów. Jednocześnie wrapper ten umożliwia nieprzerwane pobieranie danych przez teoretycznie nieskończenie długi czas ze względu na automatyczne odświeżanie tokena dostępowego zanim zostanie unieważniony. Nieprzerwane działanie takiego procesu jest kluczowe do efektywnego zebrania większych zbiorów danych, które są wymagane do efektywnego wyćwiczenia modeli klasyfikujących.

### **2.1.3 Wykorzystane endpointy**

Aby pobrać komplet interesujących nas danych wykorzystano następujące endpointy

1. <https://api.spotify.com/v1/tracks/>  
- do pobrania podstawowych danych o utworze jak identyfikator, nazwa i identyfikatory artystów, którzy są autorami danego utworu
2. <https://api.spotify.com/v1/artists/>  
- do pobrania podstawowych

danych o artystach na podstawie ich identyfikatora. Dane pobierane z tego endpointa to nazwa artysty oraz gatunki z nim skojarzone.

3. <https://api.spotify.com/v1/audio-features/> - do pobrania metadanych o utworze na podstawie jego identyfikatora. Do metadanych dostępnych pod tym endpointem należą: 'acousticness', 'danceability', 'energy', 'instrumentalness', 'key', 'loudness', 'liveness', 'mode', 'speechiness', 'tempo', 'time\_signature', 'valence'
4. <https://api.spotify.com/v1/playlists/>  
- do pobrania identyfikatorów utworów z playlisty określonego użytkownika
5. [https://api.spotify.com/v1/users/{user\\_id}/playlists](https://api.spotify.com/v1/users/{user_id}/playlists)  
- do pobrania playlist określonego użytkownika

### **2.1.4 Struktura pobranych danych**

Pobrany zbiór danych posiada pewne następujące zasady

1. Każdy utwór skojarzony jest z co najmniej jednym artystą
2. Każdy artysta może być skojarzony z dowolną liczbą gatunków
3. Każdy utwór posiada przypisany zbiór metadanych

Na podstawie powyższych zasad widać, że jedynym sposobem na powiązanie utworu z jego gatunkiem jest skorzystanie z przechodniej relacji z gatunkiem poprzez artystę. Dla pojedynczych utworów takie przypisanie gatunków może być my-

lace, ponieważ ten system zakłada, że wszystkie utwory artysty posiadają wszystkie jego gatunki, co w wielu przypadkach może być nieprawda.

#### **2.1.5 Przechowywania danych**

## **References**