Projekt IUM

Rozpoznawanie gatunków muzycznych

Mariusz Pakulski

Marcin Górski

Definicja Problemu Biznesowego

Serwis muzyczny „Pozytywka” umożliwia użytkownikom odtwarzanie ulubionych utworów online. Jednak firma napotyka problem z brakiem przypisanych gatunków muzycznych dla niektórych nowo dodawanych wykonawców. Ten brak informacji ogranicza zdolność serwisu do efektywnego katalogowania utworów, personalizacji doświadczeń użytkowników oraz wprowadzania nowych funkcji bazujących na gatunkach muzycznych.

Zadanie Modelowania

Głównym zadaniem modelowania jest opracowanie systemu klasyfikacji gatunków muzycznych dla wykonawców i ich utworów, którzy nie mają przypisanych gatunków w bazie danych. System ten powinien być w stanie automatycznie identyfikować i przypisywać odpowiednie gatunki muzyczne na podstawie dostępnych danych, takich jak:

- Charakterystyki audio utworów muzycznych,
- Metadane utworów i wykonawców,
- Opcjonalnie: *Wzorce słuchania użytkowników*.

Założenia

1. **Dostępność Danych**: Zakładamy, że:
    a) Mamy dostęp do pełnej bazy danych utworów i wykonawców (oraz opcjonalnie do historii sesji użytkowników).
    b) Charakterystyka audio utworu pozwala na jednoznaczną identyfikację gatunku.
    c) Gatunek wykonawcy to gatunek jego utworów.
    d) Artysta może mieć więcej niż jeden gatunek.
2. **Jakość Danych**: Wszystkie potrzebne dane są kompletne i aktualne.
3. **Wydajność Systemu**: System klasyfikacji musi być wydajny, aby obsłużyć duże ilości danych.

Kryteria Sukcesu

1. **Kompletność Wyników**: Przypisanie wszystkim wykonawcom gatunków muzycznych.
2. **Dokładność Klasyfikacji**: Wysoka dokładność przypisywania gatunków muzycznych do wykonawców i utworów, mierzona za pomocą metryk takich jak precyzja i czułość.

Analiza Danych

Analiza wstępna

Większość artystów ma przypisany gatunek – a właściwie listę gatunków. Oznacza to, że artysta może grać w różnych gatunkach. Dlatego możemy starać się identyfikować wszystkie z nich. Niektórzy artyści nie mają przypisanego identyfikatora, co raczej uniemożliwia powiązanie ich z utworami.

Utwory mają przypisane różnego rodzaju parametry (charakterystyka muzyczna). Nie mają jednak przypisanych gatunków, co oznacza, że w prosty sposób nie zbudujemy jedynie na podstawie tych danych klasyfikatora gatunków utworu. **W miarę możliwości prosimy o bazę utworów wraz z ich gatunkami.** To by znacząco ułatwiło sprawę i umożliwiło dokładniejszą klasyfikację wykonawców. Utwory z reguły mają bowiem przypisane id artysty (choć nie wszystkie), więc na podstawie ich gatunków można określić gatunki wykonawców. Jeśli nawet nie wszystkie utwory będą miały przypisany gatunek, to jednak odpowiednia duża liczba przypisanych gatunków pozwoli zbudować klasyfikator, który uzupełni resztę na podstawie charakterystyk dźwiękowych. Co więcej, niektóre utwory nie mają przypisanych wykonawców.

Alternatywnie możemy podejść do tematu w ten sposób:

Znamy ulubione gatunki użytkowników. Możemy prześledzić jakich utworów słuchali oni najczęściej i na tej podstawie stwierdzić, że te utwory należą do grupy ulubionych gatunków. To nie wystarczy jednak aby ustalić, które są które. Jednak jeśli dla wszystkich użytkowników tego dokonamy, otrzymamy dla danego utworu możliwe listy gatunków na podstawie różnych użytkowników. Wówczas będziemy mogli ustalić część wspólną. Może pozostanie jeden gatunek, ale może zdarzyć się tak, że pozostanie ich więcej.

Dodatkowo, znając ulubione gatunki użytkownika oraz utwory których najczęściej słucha, możemy sprawdzić do których artystów te utwory należą. Możemy ustalić w ten sposób ulubionych artystów. Część z nich będzie miała zapewne przypisane gatunki – jeśli którymś brakuje gatunku, a zarazem nie znaleźliśmy jeszcze artystów z pozostałych ulubionych gatunków użytkownika, to możemy próbować dopasować na tej podstawie artystów do gatunków.

Takie podejście jest jednak skomplikowane i trudno przewidzieć rezultaty. Dlatego ponownie, **w miarę możliwości prosimy o bazę utworów wraz z ich gatunkami** – taka informacja może pomóc znacząco podnieść dokładność klasyfikacji. W ostateczności, możemy jednak przyjąć następująca podejście:

1) wybrać wykonawców, którzy mają określony dokładnie jeden gatunek (i nadany identyfikator)

2) na tej podstawie przypisać gatunek ich utworom.

3) zbudować klasyfikator gatunku utworu na podstawie charakterystyki muzycznej

4) dokonać klasyfikacji utworów za pomocą w/w klasyfikatora

5) uzupełnić gatunki wykonawców na podstawie prognozowanych (odgadniętych) gatunków ich utworów

Dla dokładności kluczowe jest tutaj to, aby artystów z punktu 1 było wystarczająco dużo – by w ten sposób uzyskać wiele utworów, z jednoznacznie przypisanymi gatunkami. Artystów z dokładnie jednym gatunkiem jest względnie niewielu, dlatego trudno na tym etapie zagwarantować jakość takiego klasyfikatora.

Zalecenia: Dostarczenie jak najbardziej kompletnych danych; w szczególności bazy utworów uzupełnionej o gatunek muzyczny (w ostateczności o kilka gatunków muzycznych jeśli do jednego utworu może być przypisanych kilka)

Po konsultacji z klientem:

W ramach konsultacji z klientem omówiliśmy różne możliwości rozwiązania problemu. Okazało się, że niestety klient nie dysponuje ani danymi określającymi gatunek choćby części utworów muzycznych, ani danymi określającymi cechy gatunków. Dlatego zdecydowaliśmy się na następujące rozwiązanie:

1. Skorzystanie ze zbioru danych **Sporify Tracks Genre** z biblioteki Kaggle: https://www.kaggle.com/datasets/thedevastator/spotify-tracks-genre-dataset
   Zgodnie ze specyfikacją, zbiór ten dostępny jest w ramach licencji CC0: Public Domain , co umożliwia jego komercyjne zastosowanie. W zbiorze tym mamy dane dla prawie 90 tys. utworów (**89 741**). Dla każdego utworu mamy podane jego parametry **oraz** określony gatunek (**track_genre** - spośród 114 możliwych ogólnych gatunków). Cechy utworu, potencjalnie pozwalające przewidywać gatunek to:

   - **popularity** – popularność w Spotify, w zakresie od 0 do 100
   - **duration_ms** – długość trwania w ms
   - **explicit –** wartość logiczna, określająca czy utwór zawiera „explicit content”
   - **danceability –** jak bardzo utwór nadaje się do tańca (od 0 do 1)
   - **energy –** miara intensywności utworu (od 0 do 1)
   - **key** – klucz utworu
   - **loudness** – głośność utworu w dB
   - **mode –** modalność (major to 1, minor to 0)
   - **speechiness** - poziom obecności słów (od 0 do 1)
   - **acousticness –** poziom akustyczności (od 0 do 1)
   - **instrumentalness** – prawdopodobieństwo tego, że utwór jest instrumentalny (od 0 do 1)
   - **liveness** – poziom obecności publiczności podczas nagrania (na ile jest to wersja „live” – od 0 do 1)
   - **valence** – pozytywność muzyczna (radosność utworu; od 0 do 1)
   - **tempo –** tempo utworu w BPM (uderzenia na minutę)
   - **time_signature** –liczba uderzeń w każdym takcie ściezki

   Są to dokładnie te same cechy co w naszym zbiorze danych.

2. Na podstawie tego zbioru zamierzmy wytrenować klasyfikator gatunku na podstawie cech danego utworu muzycznego (uczenie nadzorowane, ponieważ dysponujemy właściwymi etykietami klas).
3. Wytrenowany klasyfikator zamierzamy użyć do przypisania gatunków utworom z naszej bazy danych.
4. Na podstawie gatunków utworów, zamierzmy przypisać gatunki ich wykonawcom.

Uwzględnienie uwag:

Modelowanie

1. Jak porównuje się zaproponowane kryterium sukcesu do modelu naiwnego (zwracającego zawsze taki sam wynik)?

   Model naiwny zwracałby zawsze najczęściej występujący gatunek muzyczny, uzyskując dokładność równą częstości najczęstszej klasy.
   Naszym celem jest uzyskanie istotnie wyższych wartości metryk (precyzji i czułości) niż model naiwny.

2. Ustalone kryteria sukcesu powinny być jakoś oparte o dostępne dane.

   Stwierdziliśmy, że naszym kryterium sukcesu jest "Dokładność Klasyfikacji: Wysoka dokładność przypisywania gatunków muzycznych do wykonawców i utworów, mierzona za pomocą metryk takich jak precyzja i czułość."

   W oparciu o dostępne dane, zamierzamy porównać wyniki naszego modelu z wynikami modelu naiwnego. Za sukces uznamy, jeśli nasze wyniki okażą się istotnie wyższe.

3. Jakie jest biznesowe kryterium sukcesu?

   Chcąc ustalić biznesowe kryterium sukcesu, zadajemy sobie pytania pomocnicze – co chcemy zrobić?

   - Poprawiamy istniejące podejście (niekoniecznie stosujące UM)?
   - Spełniamy jasno sprecyzowane wymaganie klienta?
   - Staramy się przewyższyć konkurencję/standardy branżowe?
   - Prowadzimy wstępne prace badawcze/"rozpoznajemy temat" (z ograniczeniami czasowymi/zasobowymi/budżetowymi)?
   - Klienta interesuje zwrot z inwestycji?

   W tym przypadku mamy do czynienia z jasno sprecyzowanymi wymaganiami klienta:

   "Nie wszyscy nowo dodawani wykonawcy do naszej bazy mają przypisany gatunek

   muzyczny – musimy jakoś temu zaradzić!"

   Klient zatem jasno określił co jest naszym celem – przypisanie gatunku muzycznego wszystkim wykonawcom. Oczywiście zgodnie z naszym podejściem jest to możliwe dla wykonawców, do których mamy przypisany przynajmniej jeden utwór w dostarczonych danych.

A zatem nasze biznesowe kryterium sukcesu możemy sformułować następująco:

Każdy wykonawca, do którego mamy przypisany przynajmniej jeden utwór w dostarczonych danych, ma mieć przypisany co najmniej jeden gatunek.

Oczywiście gatunek ten ma zostać przydzielony w miarę możliwości zgodnie z prawdą, a żeby to ocenić zastosujemy **analityczne** kryterium sukcesu…

4. Jakie jest analityczne kryterium sukcesu?

   Precyzja i czułość mają mieć wartości istotnie wyższe niż dla modelu naiwnego (dla danych ze zbioru Spotify)

   W celu określenia czy to kryterium zostało spełnione, podzielimy zbiór danych na część uczącą, testową i walidacyjną. Na zbiorze uczącym będziemy uczyć model, na walidacyjnym będziemy dobierać hiperparametry, a testowy posłuży do określenia czy udało nam się spełnić kryterium sukcesu. Wykorzystamy zbiór danych Spotify, ponieważ posiada on przypisanie gatunków do utworów, co umożliwi zarówno naukę, jak i weryfikację.

5. Jaka jest akceptowalna wartość proponowanej metryki w analitycznym kryterium sukcesu?

   Akceptowalna wartość ustalona zostanie na podstawie wartości dla modelu naiwnego.

   a) Precyzja: precision = TP/(TP+FP)

      Czyli liczba utworów prawidłowo sklasyfikowanych jako dany gatunek, w stosunku do liczby wszystkich utworów sklasyfikowanych jako dany gatunek. Innymi słowy: jaka część utworów sklasyfikowanych jako dany gatunek, rzeczywiście ma taki gatunek. Należy tę wartość wyznaczyć dla danego gatunku osobno, a na koniec te wyniki uśrednić.

      W naszym zbiorze Spotify mamy 114 gatunków i każdy gatunek występuje dokładnie 1000 razy (zbiór jest doskonale zbalansowany). Oznacza to, że każdy z nich jest najczęściej występujący i nasz naiwny klasyfikator mógłby wybrać dowolny z nich jako ten, na który zawsze by wskazywał. Dla ustalenia uwagi przyjmijmy, że byłby to gatunek **acoustic**.

      Naiwny klasyfikator twierdziłby, że każdy utwór jest gatunku acoustic. W przypadku 1 tysiąca utworów byłaby to prawda (TP = 1000), a w przypadku 113 tysięcy fałsz (FP = 113 000). Uzyskana precyzja dla gatunku acoustic wyniosłaby zatem: 1000/114000 ~= 0,0088, czyli 0,88% .Dla pozostałych gatunków nie mielibyśmy żadnych „obstawień" (dzielilibyśmy 0 przez 0), więc należy je pominąć przy liczeniu średniej. Średnia precyzja dla klasyfikatora naiwnego wynosiłaby zatem 0,88%.
      Ponieważ jest to niski wynik, jako akceptowalną wartość precyzji w naszym analitycznym kryterium sukcesu proponujemy 4-krotność tej wartości, czyli min. **3.52%**. Oznacza to, że za sukces uznamy jeśli nasz model uzyska co najmniej 4 razy większą średnią precyzję niż model naiwny.

b) Czułość: recall = TP/(TP+FN)

Czyli liczba utworów prawidłowo sklasyfikowanych jako dany gatunek, w stosunku do liczby wszystkich utworów danego gatunku. Innymi słowy: jaką część utworów danego gatunku udało nam się wykryć. Należy tę wartość wyznaczyć dla danego gatunku osobno, a na koniec te wyniki uśrednić.

Naiwny klasyfikator twierdziłby, że każdy utwór jest gatunku acoustic. Wykryłby zatem w 100% każdy utwór gatunku acoustic. Jednak dla pozostałych utworów nie wykryłby ani jednego utworu. Średnia czułość wyniosłaby zatem 100%/114 ~= 0,88%

Ponieważ jest to niski wynik, jako akceptowalną wartość precyzji w naszym analitycznym kryterium sukcesu ostrożnie proponujemy 4-krotność tej wartości, czyli min. **3.52%**. Oznacza to, że za sukces uznamy jeśli nasz model uzyska co najmniej 4 razy większą średnią czułość niż model naiwny.

Podsumowując: nasze analityczne kryteria sukcesu to uzyskanie średniej precyzji na poziomie co najmniej 3.52% oraz średniej czułości na poziomie także co najmniej 3.52%.

Eksploracyjna analiza danych

1. Czy w danych występują błędne/brakujące atrybuty?

W przypadku danych ze Spotify (na których będzie budowany klasyfikator) nie stwierdzono błędnych/brakujących atrybutów. Jednak w przypadku danych otrzymanych od klienta stwierdzono braku w następujących atrybutach (spośród omówionych wcześniej atrybutów istotnych dla tego zadania): **popularity** oraz **mode**. Dlatego zdecydowano, że nie będziemy na nich polegać przy rozpoznawaniu gatunku. Zostaną one zatem pominięte w dalszych analizach.

Do dalszych rozważań pozostają nam zatem (póki co) następujące atrybuty wejściowe:

- **duration_ms** – długość trwania w ms
- **explicit –** wartość logiczna, określająca czy utwór zawiera „explicit content”
- **danceability –** jak bardzo utwór nadaje się do tańca (od 0 do 1)
- **energy –** miara intensywności utworu (od 0 do 1)
- **key** – klucz utworu
- **loudness** – głośność utworu w dB
- **speechiness** - poziom obecności słów (od 0 do 1)
- **acousticness –** poziom akustyczności (od 0 do 1)
- **instrumentalness** – prawdopodobieństwo tego, że utwór jest instrumentalny (od 0 do 1)
- **liveness** – poziom obecności publiczności podczas nagrania (na ile jest to wersja „live” – od 0 do 1)
- **valence** – pozytywność muzyczna (radosność utworu; od 0 do 1)
- **tempo –** tempo utworu w BPM (uderzenia na minutę)
- **time_signature** –liczba uderzeń w każdym takcie ściezki

2. Czy jesteście Państwo pewni, że zmienne wejściowe niosą jakąś informację o zmiennej celu?

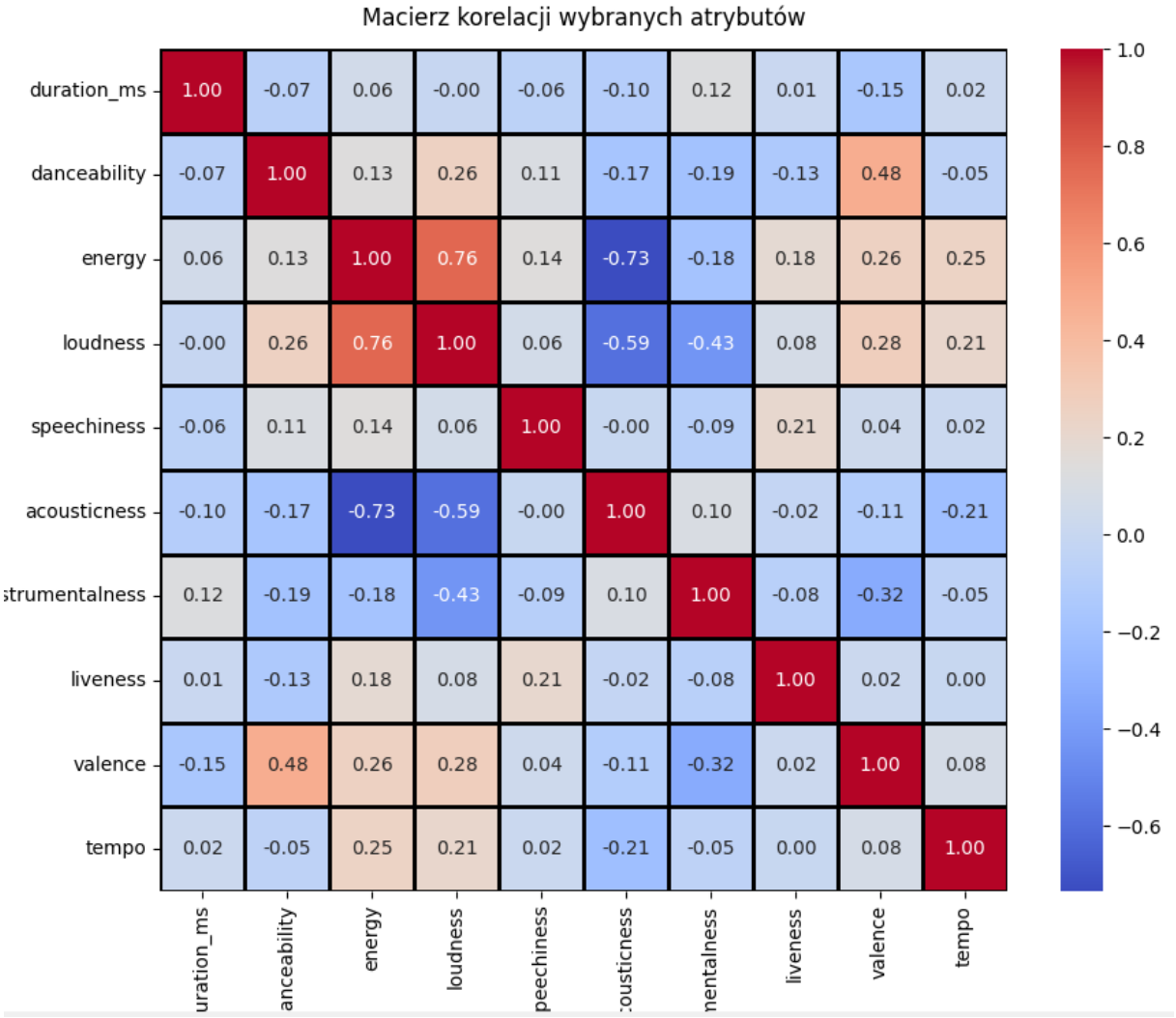
   Zmienna celu jest atrybutem dyskretnym nominalnym, więc nie możemy w celu ustalenia tego stosować współczynników korelacji (ani liniowej ani rangowej). Wykrywają one bowiem jedynie zależności liniowe lub monotoniczne. Dlatego w naszej sytuacji posłużymy się współczynnikiem informacji wzajemnej. Gatunki dla utworów mamy podane jedynie w zbiorze ze Spotify, więc to dla niego wykonamy obliczenia. Ponieważ część (a właściwie większość) zmienncyh wejśćiowych ma charakter ciągły, zastosujemy do obliczeń metodę *mutual_info_classif* z sklearn.feature_selection (ze wskazaniem, które zmienne są dyskretne, a które ciągłe), która estymuje wartość informacji wzajemnej (bez potrzeby dyskretyzacji). Oto otrzymane wyniki:

Atrybut	Informacja Wzajemna
`acousticness`	0.630
`tempo`	0.595
`loudness`	0.539
`duration_ms`	0.528
`energy`	0.495
`instrumentalness`	0.412
`danceability`	0.410
`valence`	0.362
`speechiness`	0.322
`liveness`	0.221
`explicit`	0.066
`time_signature`	0.054
`key`	0.041

   Najwięcej informacji o zmiennej celu niosą zatem następujące zmienne wejściowe: **acousticness, tempo, loudness, duration_ms, energy, instrumentalness** i **danceability.** Pewną wartość informacyjną mają także zmienne **valence, speechiness** i **liveness.** Natomiast najmniej informatywne okazały się zmienne: **explicit, time_signature** oraz **key**. Dlatego też zostaną one pominięte. Pozostają zatem:

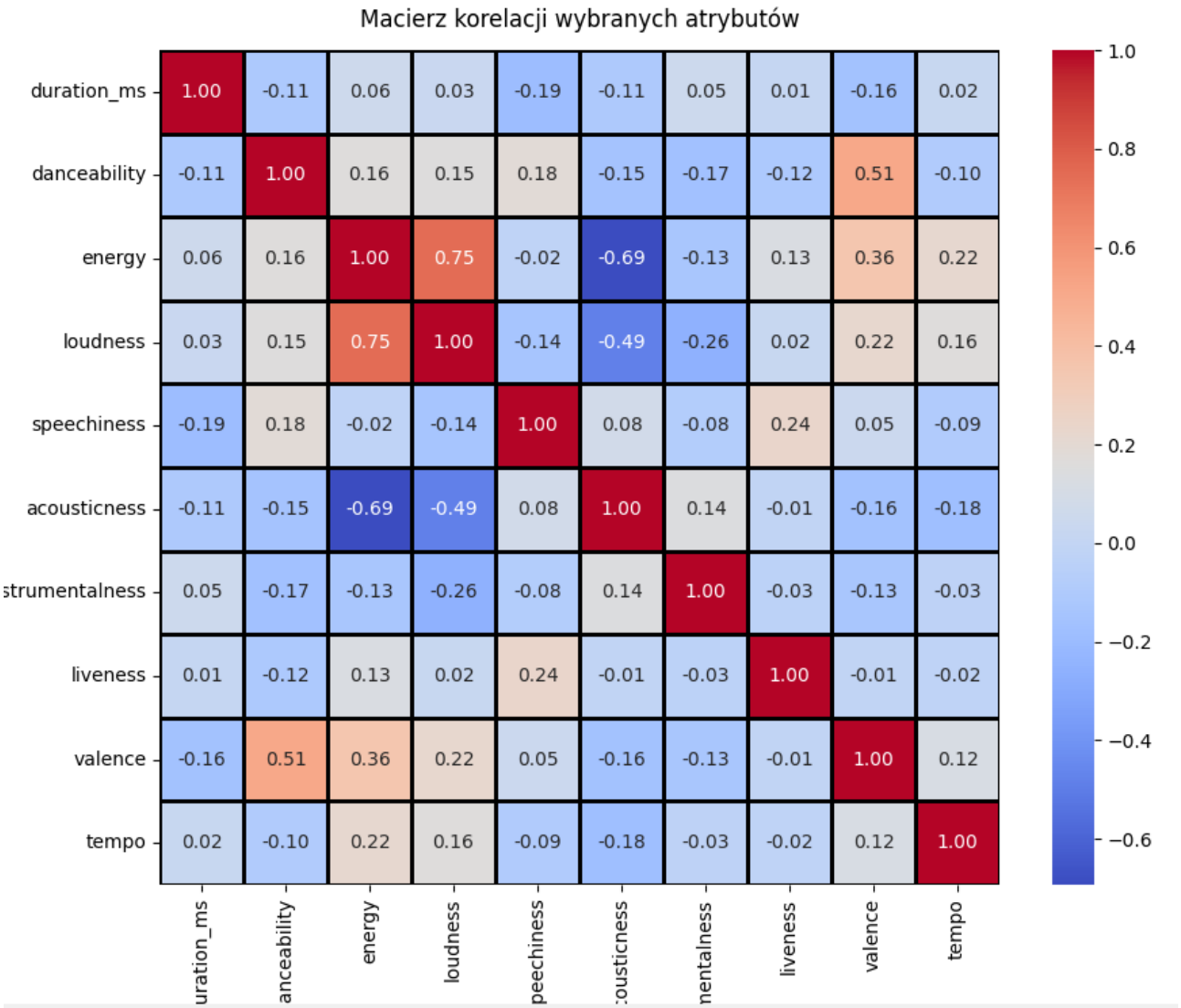
- **duration_ms** – długość trwania w ms
- **danceability –** jak bardzo utwór nadaje się do tańca (od 0 do 1)
- **energy –** miara intensywności utworu (od 0 do 1)
- **loudness** – głośność utworu w dB
- **speechiness** - poziom obecności słów (od 0 do 1)
- **acousticness –** poziom akustyczności (od 0 do 1)
- **instrumentalness** – prawdopodobieństwo tego, że utwór jest instrumentalny (od 0 do 1)
- **liveness** – poziom obecności publiczności podczas nagrania (na ile jest to wersja „live" – od 0 do 1)
- **valence** – pozytywność muzyczna (radosność utworu; od 0 do 1)
- **tempo –** tempo utworu w BPM (uderzenia na minutę)

Na placu boju pozostały zatem same zmienne ciągłe. Możemy więc stworzyć dla nich macierz korelacji liniowych (dla danych ze Spotify):



acousticness, energy i **loudness** są dość mocno skorelowane ze sobą (akustyczność jest ujemnie skorelowana z energią i głośnością, a energia i głośność dodatnio ze sobą). Nie są to jednak wartości na tyle duże, by już na tym etapie wykluczać któreś z nich. Pewną umiarkowanie wysoką korelację wykazują także pary **danceability** i **valence** oraz **instrumentalness** i **loudness**. Pozostałe zmienne wykazują niskie korelacje z innymi atrybutami.

Podobną macierz możemy stworzyć dla danych otrzymanych od klienta:

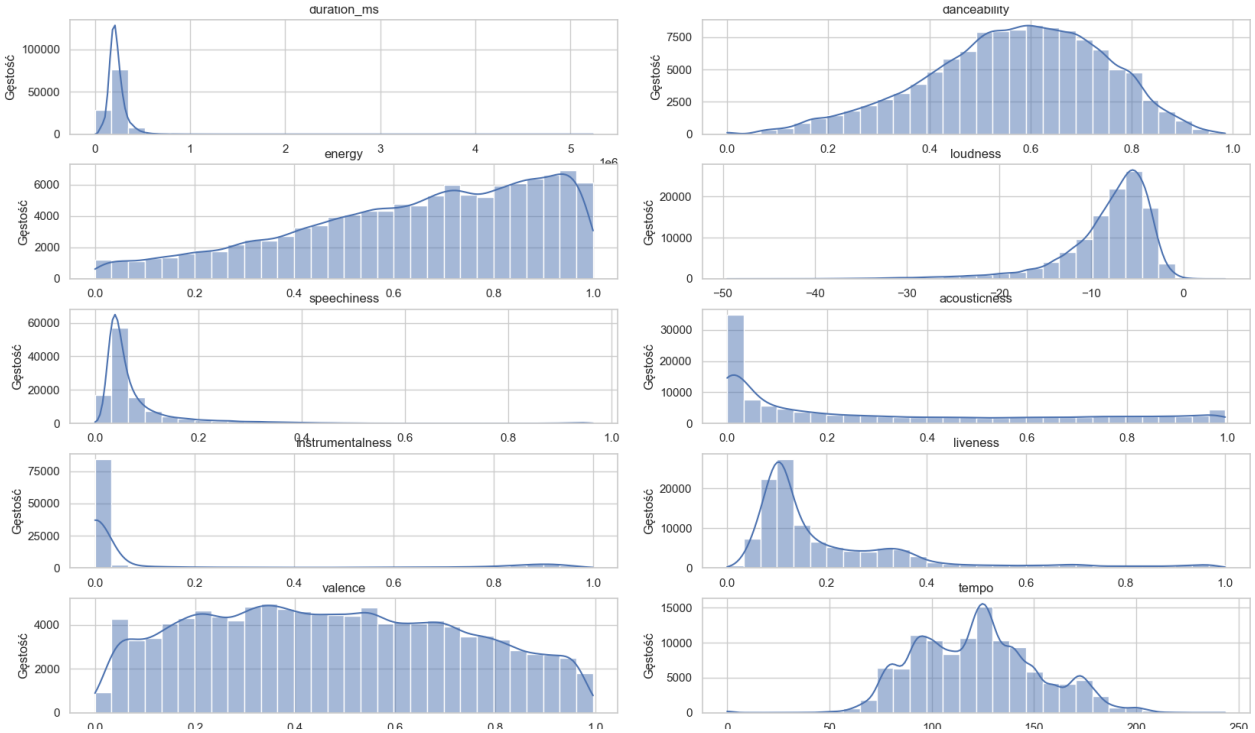


Widzimy, że wyniki są dość podobne. Jedynie znacząco słabnie korelacja między **instrumentalness** i **loudness** oraz dochodzi dodatkowo umiarkowana korelacja między **danceability** i **valence**.

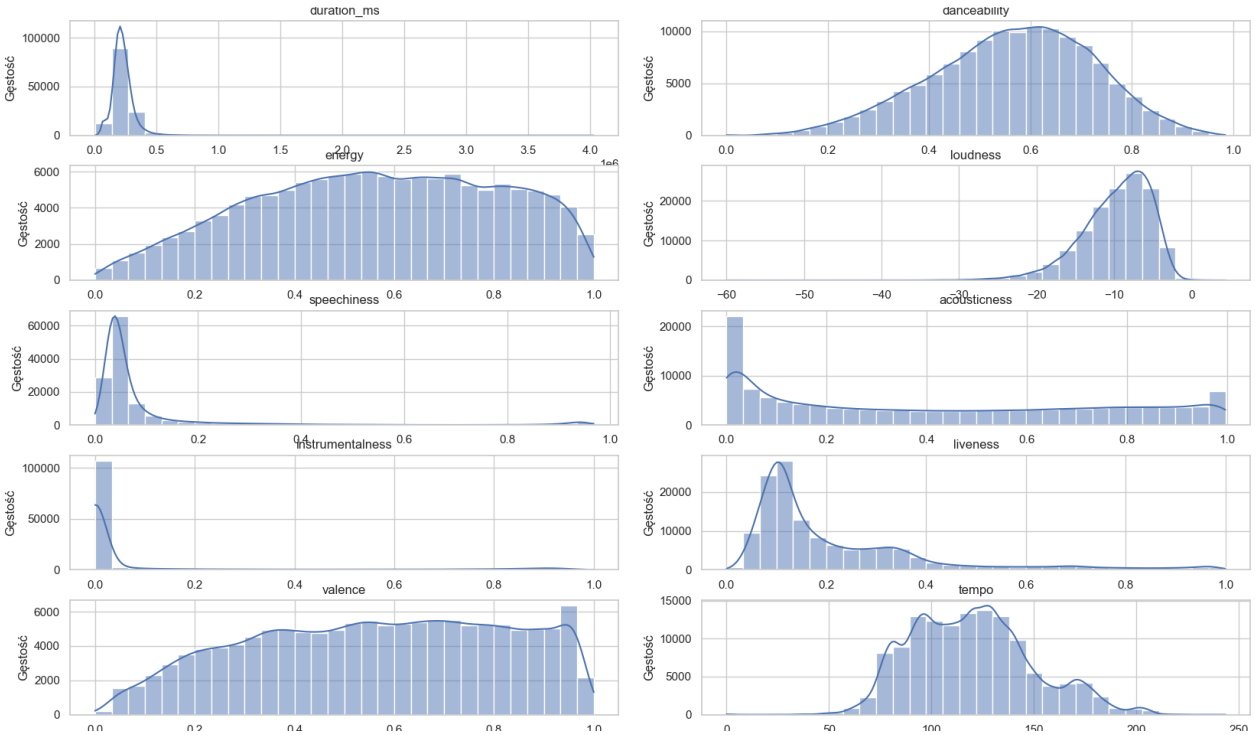
3. Jak wyglądają rozkłady kluczowych do realizacji projektu atrybutów?

Dla danych ze Spotify mamy 114 tysięcy utworów w 114 gatunkach – każdy z nich występuje 1000-krotnie. Dane od klienta pozbawione są informacji o gatunku. Co się zaś tyczy atrybutów wejściowych (linią ciągłą zaznaczone estymowane rozkłady):

a) Dla danych ze Spotify:



b) Dla danych otrzymanych od klienta:



Jak widzimy, rozkłady te są podobne w obu przypadkach. Wartości z niektórych zakresów występują bardzo rzadko. Jeśli skupimy się jednak na najgęstszych obszarach wykresów, to zauważymy, że rozkłady dla **danceability** i **loudness** przypominają rozkłady normalne. **liveness** i **speechiness** też można uznać za normalne (prawoskośne). Rozkład dla **acousticness** przypomina natomiast rozkład jednostajny. **instrumentalness** najczęściej ma wartość bliską 0 (zdecydowana większość utworów ma bardzo niskie prawdopodobieństwo bycia instrumentalnymi).

duration_ms ma rozkład prawoskośny z długim ogonem w kierunku większych wartości. Oznacza to, że większość utworów ma krótki czas trwania (najczęściej nie przekracza 0.25*1e6 ms czyli ok. 4 minut), ale istnieje trochę znacznie dłuższych.

valence ma rozkład zbliżony do jednostajnego, ale z lekką skośnością – dla danych Spotify lekka prawoskośność (przewaga utworów o mniejszej pozytywności), a dla danych od klienta lekka lewoskośność (przewaga utworów o większej pozytywności).

Rozkład dla **tempo** dla Spotify wygląda mniej więcej na multimodalny (kilka szczytów, zapewne odpowiadających bardziej popularnym tempom w muzyce). W przypadku danych od klienta podobnie, choć wykres jest bardziej wygładzony, przez co bardziej przypomina rozkład normalny.

energy ma w przypadku Spotify rozkład liniowo rosnący – im wyższy poziom energii, tym częściej występuje. W przypadku danych od klienta otrzymujemy rozkład przypominający spłaszczony rozkład normalny (z pewną lewoskośnością).

Na tym etapie nie widzimy podstaw by kierując się estymowanymi rozkładami wykluczyć kolejne zmienne wejściowe. Rzadko występujące wartości mogą okazać się istotne dla niektórych z naszych wielu (114) gatunków.

Podobieństwo zbioru rozkładów sugeruje, że wyniki modelu, uzyskującego wysokie wartości metryk dla danych ze Spotify, można potraktować jako wiarygodne dla serwisu muzycznego klienta.