Politechnika Wrocławska
Wydział Informatyki i Telekomunikacji

Metody i Systemy Decyzyjne

Sprawozdanie z laboratorium

Autor

Aleksander Woźniak

nr albumu: **272736**

kierunek: **Informatyka Stosowana**

8 czerwca 2024

Streszczenie

Celem pracy jest stworzenie modelu uczenia maszynowego, który będzie zdolny rozpoznawać gatunki muzyczne dla wprowadzonych piosenek, co pozwoli nam na zbadanie dziedziny muzyki z perspekrtywy Computer Science. Na tym polu wielu badaczy i inżynierów przeprowadzało już analizy i tworzyło swoje modele oparte zarówno o CNN (Convolutional Neural Networks)[3] jak i metody bazujące na ekstrakcji cech[2]. Niniejsza praca zaprezentuje autorskie podejście do problemu, stosując motyw ekstrakcji cech z datasetu, pobranego ze źródeł, oferujących playlisty posegregowane gantunkami. Zestaw ekstraktowanych cech będzie składał się zarówno ze standardowych cech określanych dla tego problemu, jak i z autorskich, opracowanych na potrzeby zadania. Dla przygotowanego datasetu przetestujemy kilka typów modeli operujących na danych cechach i wybierzemy ten, który da najlepszy wskaźnik *accuracy*. Pytaniami na które spróbuje odpowiedzieć praca będą: Jakie są najbardziej znaczące cechy determinujące gatunek. Na bazie modelu, jakie gatunki są do siebie podobne, Jak skategoryzować bardziej złożone gatunki na bazie modelu.

1 Wstęp – sformułowanie problemu

Dla ludzkiego ucha problem rozpoznania gatunku czesto bywa trywialny. Może czasami mamy trudność z klasyfikacją jakiejś piosenki jeśli nie oriętujemy się w danym stylu muzycznym lub piosenka jest nietuzinkowa. W kontrapunkcie, dla komputera piosenka jest niczym innym jak falą dźwiękową zapisaną w danym formacie próbek na sekundę i rozdzielczości bitowej. Patrząc na problem z tej perspektywy na pierwszy rzut oka może wydawać się on nie do rozwiązania. W końcu jak prawie metafizyczne odczucie muzyki, które odczuwamy wsłuchując się w jej kunszt może być zastąpione przez obliczenia bazujące na wykresie falowym. W celu zawężenia zakresu pracy skupimy się na klasyfikacji wymienionych gatunków: *Muzyka Klasyczna, Electro, Metal, Jazz, Funk, Progresywny Metal, Rap, Punk Rock* Końcowym celem przedsięwzięcia jest wykorzystanie modelu komputerowego do zrozumienia ukrytych cech muzyki. Cechy te będziemy badać zgodnie z następującymi etapami:

- Sprawdzimy, które cechy są najważniejsze w predykcji
- Wykorzystamy wytworzony przez nas model jako narzędzie poznania ukrytych powiązań między gatunkami
- Dla wytworzonego modelu sprawdzimy również klasyfikację piosenek zespołu *Queens of the Stone Age*, posiadającego spory zasób różnorodnej muzyki, z której większość bywa klasyfikowana jako psychodeliczny rock.

2 Opis rozwiązania

Rozwiązaniem tego problemu w tym wydaniu jest wyekstraktowanie z każdej piosenki zbioru cech, licząc na to, że z wysokim prawdopodobieństwem zbiór wartości tych cech definiuje ich gatunek (np. tempo 130bpm i długość 3:56min daje wieksze szanse na klasyfikacje gatunku pokroju metal czy electro, w przeciwieństwie do muzyki klasycznej)

Dane dla modelu zostaną pobrane ze strony *strona usunięta*, która umożliwia użytkownikom wolne udostępnianie treści. Na owej stronie zostaną zidentyfikowane playlisty skomponowane przez słuchaczy zoriętowane rodzajami dla szukanych przez nas gatunków. Taki "model" rozpoznawania gatunków (ludzki układ słuchowy i powiązane z nim procesy poznawcze) będziemy traktować dalej jako punkt odniesienia *Grand Truth*.

Genre	count
Blues	47
Classical	113
Elctro	75
Funk	42
Jazz	64
Metal	65
Prog Metal	38
Punk	24
QOTSA	119
Rap	100
total	687

Tabela 1: ilość piosenek w datasecie

Do konstrukcji modelu wyznaczymy cechy do ektrakcji. Do ekstrakcji cech wykorzystamy głównie moduł **librosa**. Autorskie cechy zostały skonstruowane tak, aby w razie potrzeby były znormalizowane (np. dzielenie danego wyznacznika przez długość piosenki). Dla skupienia uwagi podzielimy je na oczywiste lub przyjęte w sztuce:

- tempo - wyrażone w uderzeniach na minutę
- długość piosenki - wyrażone w sekundach
- argument maksimum spektralnej centroidy w proporcji do długości piosenki - oznacza w której części piosenki najbardziej zaakcentowaną częstotliwością jest najwyższa częstotliwość
- mediana spektralnego spadku (spectral rolloff) - mediana częstotliwości w piosence, pod którymi stężenie dźwięku zajmuje podany procent spektrum (sprawdzimy tę cechę zarówno dla wysokiej, jak i niskiej wartości granicy spadku)

oraz na takie skomponowane w toku tworzenia modelu na potrzeby ninejszej pracy:

- powtarzalność
- wariacja tempa
- waraicja głośności
- zmiana kontekstu muzycznego
- długość zestawu dźwięków występujących często
- stopień obecności perkusji
- mediana zaakcentowanych częstotliwości

Dla tak zaplanowanego działania sprawdzimy, które intuicje są prawidłowe, a które miały mankamenty lub były błędne. Należy pamiętać, że mimo celnych intucji, działania prowadzące do ekstrakcji ograniczone są przez zestaw narzędzi dostępnych w używanych modułach (np. **Librosa — melodic spectrogram, chromagram** czyli cechy pozyskane w wyniku zastosowania na danych fali

dźwiękowej Short-time Fourier transform, która rozdziela częstotliwości występujące w sygnale za pomocą Transformaty Fouriera i dalej przetwarzane w celu ekstrakcji cech). Cechy, których rozkłady w populacjach gatunkowych będą budziły wątpliwości, będą usuwane. Jeśli wszystkie cechy będą budziły wątpliwości, a wyniki klasyfikacji będą wykazywały się poprawnym działaniem zignorujemy problem, przyjmując wadę obróbki i akceptując nieidealny charakter preprocessingu danych.

3 Rezultaty obliczeń

3.1 Plan badań

Aby wytrenować model, musimy najpierw skonstruować zestaw cech. Z takiego zestawu będziemy w stanie ocenić które cechy ujawniają zakłócenia, co pozwoli nam podjąć decyzję co do dalszego przerabiania naszego datasetu. Wyniki wytrenowanego modelu po usunięciu danych sprawiających trudności powinny działać lepiej niż modelu wytrenowanego na całym datasecie. Należy pamiętać jednak, że zbiór testowy zapewne cierpi na te same mankamenty co zbiór treningowy. Jednocześnie warto pamiętać, że autor pracy nie dysponuje zasobami pozwalającymi na przeprowadzenie badania w sposób perfekcyjny (przez sam problem pozyskania datasetu, który jest obarczony pewną dozą stronniczości próbkowania oraz jest ograniczony) i trzeba liczyć się z pilotarzowym, poglądowym charakterem wyników, uzyskanych w toku przeprowadzania badania.

Działania pojęte w celu rpzeprowadzenia badania możemy rozłożyć na kluczowe etapy:

- Ekstrakcja cech z pobranego datasetu piosenek formatu .mp3 i .flac

    1. Do ekstrakcji każdej z cech, będziemy ładować najpierw piosenkę do formatu fali
    2. Cechy oczywiste lub przyjęte w sztuce zostaną wyekstraktowane w standardowy sposób
    3. Powtarzalność będzie wyekstraktowana przy pomocy **Librosa — Beat track, Chromagram**. Dla pozyskanych z **Beat Track** indeksów ćwierćnut uzyskamy zestaw zagrywek rozdzielonych ośmioma ćwierćnutami. Następnie posłużymy się klasteryzacją, aby połączyć podobne zagrywki. Funkcją podobieństwa będzie ilość tych samych nut na każdej pozycji, a granicę podobieństwa ustawimy na 0.7. Intuicja jest prosta, im mniej na ilość zagrywek mamy unikalnych zagrywek, tym bardziej powtarzalna jest piosenka. Warto zauważyć, że dla metrum o nieparzystej liczbie wybranej podstawy (np. 3/4, 7/8, 5/4 etc.) nasz wynik będzie przekłamany, ponieważ okno wybierania zagrywek bedzie zahaczało o następny takt, co w rezultacie może dobrać kilka nowych zagrywek, gdy tak naprawde powinno wybrać mniej. Jednak korzystając z intuicji stwierdzimy, że piosenki o metrum innym niż 4/4 mają tendencję do przełamywania powtarzalności i kodują rozwiązania muzyczne mniej *mondain*.
    4. Wariacja tempa zostanie obliczona na podstawie próbkowania uderzeń na minutę w różnych częściach piosenki. Z takich próbek następnie zostanie obliczone odchylenie standardowe. Ta cecha ma dać nam wyznacznik zmienności tempa w piosence.
    5. Wariacja głośności policzona z odchylenia standardowego średnich głośności próbek policzonych z **Librosa — RSM** będzie mówiła nam o dynamice głośności piosenki.
    6. Zmiana kontekstu muzycznego bedzie policzona z częstości zmiany zestawu siedmiu dźwięków w piosence
    7. Długość zestawu dźwięków występujących często będzie działała na zasadzie zebrania ilości wystąpień dźwięków w piosence. Im bardziej równy bedzie rozkład owych 12 dźwięków,

tym bardziej wnioskujemy zróżnicowanie melodyczne . Ta cecha zakoduje różnorodność piosenki.

8. Stopień obecności perkusji wywnioskujemy dzięki podzieleniu fali na melodyczną i perkusyjną korzystając z **Librosa — HPSS**, który działa za pomocą filtrowania spektrum. Wyciągając mediany z obydwu części spectrum i dzieląc je na siebie uzyskamy ich proporcję.
9. Mediana zaakcentowanych częstotliwości będzie obliczona z **Librosa — Mel Spectrogram** dla którego policzymy zbió najbardziej zaakcentowanych częstotliwości, z którego wyciągniemy medianę.

- Analiza wyników ekstrakcji cech

  1. Dla każdej klasy sworzymy histogramy reprezentujące rozkłady cech w ich populacjach.
  2. Analiza rozkładów umożliwi nam stwierdzenie czy intuicje, którymi się kierowaliśmy oraz konkretne działania, które dla nich wybraliśmy były słuszne czy nie.
  3. Wyciągnięcie wniosków pozwoli nam na możliwe poprawki datasetu oraz na tworzenie dokładniejszych modeli w przyszłości.

- Stworzenie modelu przy pomocy modułu **Sci Kit Learn**

  1. Podzielimy dataset na treningowy oraz testowy tak, aby każda klasa miała w nich równy procentowy udział. Wykorzystamy do tego **Sci Kit Learn — Stratified Shuffle Split** zoretowany na gatunek.
  2. Wybierzemy najlepiej przybliżający model z pośród: **Sci Kit Learn — Gradient Boosting Classifier, Random Forest Classifier, SVC** pamiętając o przetestowaniu dla każdego z nich różnych hiperparametrów.

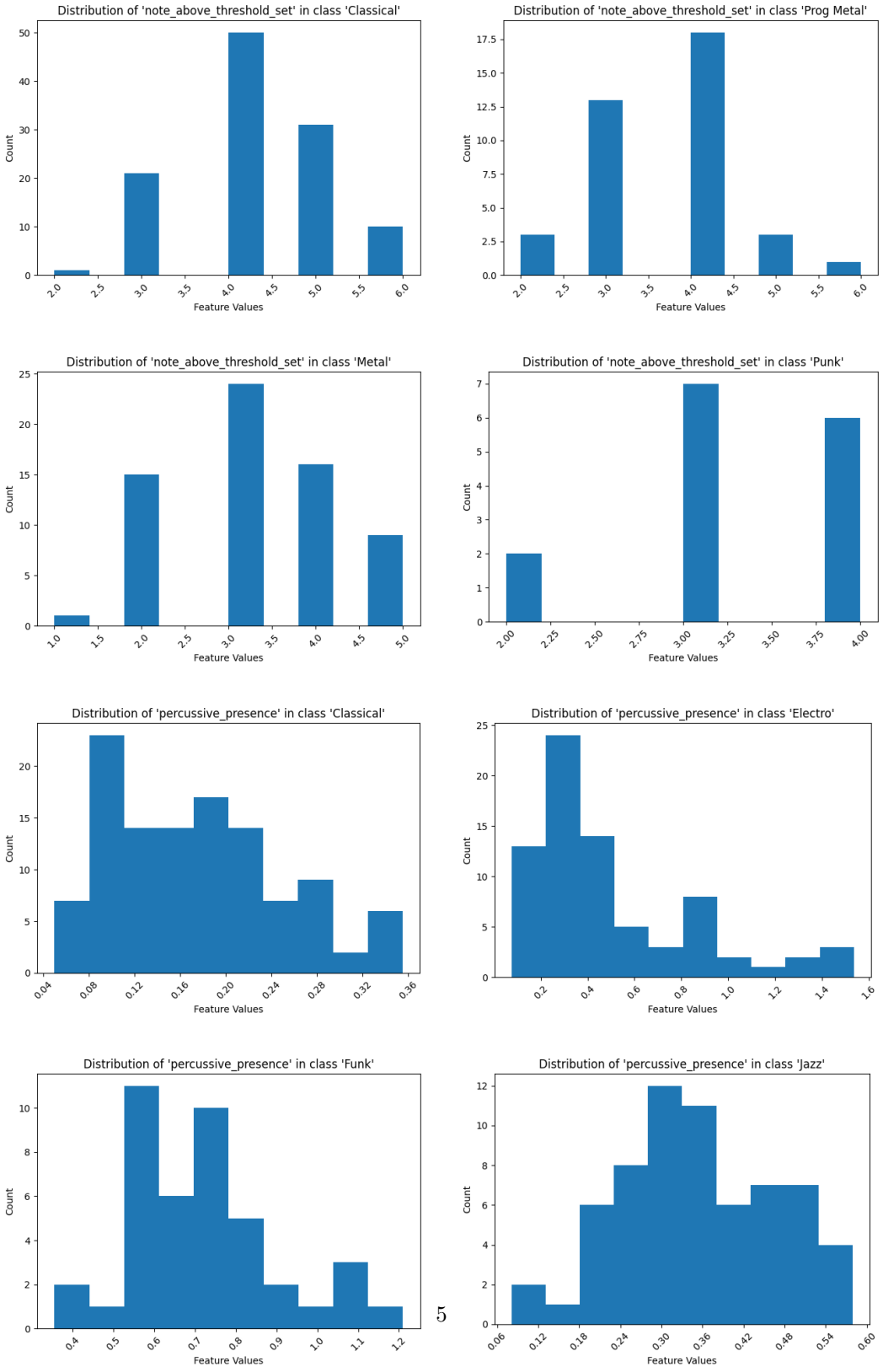
- Zbadanie odpowiedzi jakie da nam model na pytania: Jakie są najbardziej znaczące cechy determinujące gatunek. Na bazie modelu, jakie gatunki są do siebie podobne. Jak skategoryzować bardziej złożone gatunki na bazie modelu

  1. Po stworzeniu wszystkich głównych zasobów potrzebnych do wykonania zadania będziemy manipulować (odłączać i przyłączać cechy), żeby sprawdzić, której rozłączenie stanowi największe straty dla celności modelu. Da nam to pewne pojęcie o istotności każdej z cech
  2. Wytrenowany przez nas model będzie również predykował dla wprowadzonych piosenek niepoprawne gatunki. Sprawdzimy zatem, które gantunki przydziela nasz model co potraktujemy jako informację, że piosenka jest bliska innemu niż rzeczywisty gatunek, co zwrotnie potraktujemy jako podobieństwo gatunków.
  3. Po wytrenowaniu modelu sprawdzimy jak zakwalifikuje on piosenki zespołu *Queens of the Stone Age* i wyciągniemy z tego wnioski

3.2 Wyniki obliczeń

3.2.1 Ekstrakcja cech

Przeprowadzone obliczenia cech zwizualizuowane w postaci histogramów 1ujawniają nam, że wyekstraktowane cechy są raczej dobre. Świadczyć o tym mogą rozkłady wartości niektórych

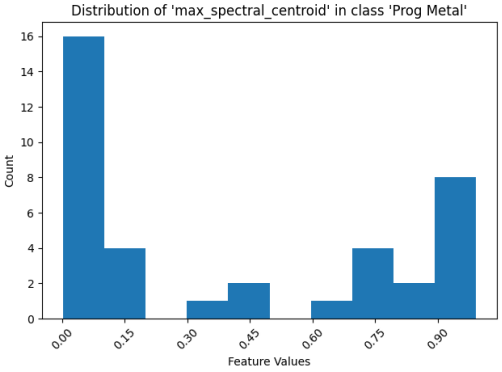


Rysunek 1: Cechy z rozkładem zbliżonym do normalnego, o różnych nadziejach i odchyleniach standardowych (prawdopodobnie dobra cecha).

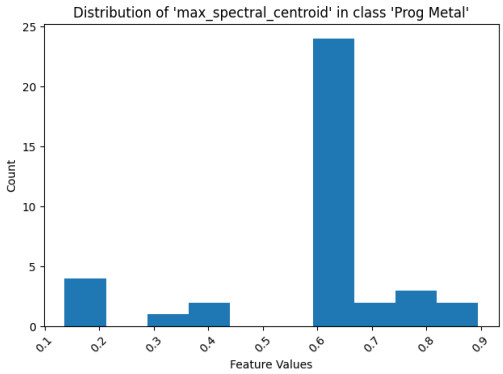
cech w poszczególnych gatunkach skoncentrowanych wokół pewnych wartości średnich zbliżone w do rozkładów normalnych, co było spodziewane, biorąc pod uwagę naturę niektórych z nich.

3.2.2 Przetwarzanie cech

Jedyna cecha, której histogramy sugerują zakłócenia jest argument maksimum spektralnej centroidy w proporcji do długości piosenki. Na wykresach możemy znaleźć wartości dla nich w skrajnych częściach przestrzeni, co oznacza, że ta cecha została obrana dla rozpoczęcia piosenki lub jego zakończenia. Sugeruje to mocno na błąd w ekstrakcji cechy. Żeby sobie z tym poradzić zimputujemy wartości nierealistycznie skrajne medianą wartości poszczególnych gatunków. Postępowanie to przeprowadzimy w celu zimputowania wartości cech outlierów z nadzieją na lepszą generalizacje modelu.



[Nie zimputowane]



[Zimputowane]

Rysunek 2: Wizualizacja imputowania przykładowej cechy

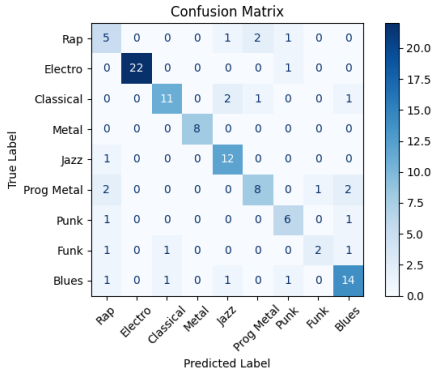
Tak zimputowana cecha została "wyczyszczona" z zakłóceń niepoprawnej ekstrakcji i przygotowana do dalszej obróbki. Średni wynik *accuracy* dla tych samych ziaren losowości przed zimputowaniem jest równy **0.74** a po **0.79**.

3.2.3 Dobranie najlepszego modelu

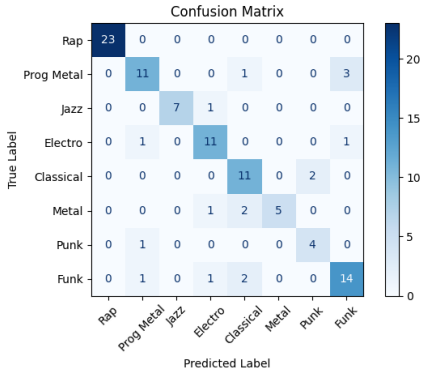
Rostrzyganie najlepszego modelu odbyło się poprzez porównanie średniej wartości *accuracy* kilku z zaproponowanych gotowych modeli. Wyniki były następujące:

- SVC — 0.55
- Random Forest Classifier — 0.65
- **Gradient Boosting Classifier** — 0.7

Z powyższych wybierzemy ten, który sprawdził się najlepiej, czyli **Gradient Boosting Classifier** (wytłumaczenie działania modelu znajduje się w bibliografii [1]. Działanie naszego modelu zobrazujemy za pomocą macierzy pomyłek:



Z danych zwrotnych wynika, że **Blues** znacznie zakłóca działanie modelu. Może to wynikać z niepoprawnego próbkowania danych tej kategorii. Odrzucimy zatem tę klasę przez co będziemy mieli mniejszy zakres działania, ale za to celniejszy model. Macierz pomyłek bez **Blues**:



3.2.4 Weryfikacja cech

Mimo, że we wcześniejszym punkcie poprawialiśmy niektóre cechy to nadal nie mamy pewności jak każda cecha wpływa na nasz model (niektóre cechy mogą wręcz oznaczać zakłócenia dla naszego mo-

delu). Przechodząc zatem do tematu dobierania cech wprowadzimy szybki *sanity check* sprawdzając jak rozłączenie każdej z nich wpływa na *accuracy*:

feature dropped	**delta accuracy**
tempo_variation	-0.0054
BPM	-0.0098
repetitiveness	0.0045
seconds_duration	-0.0241
loudness_variation	-0.0071
max_spectral_centroid	-0.0643
median_spectral_rolloff_high_pitch	-0.0054
median_spectral_rolloff_low_pitch	-0.0125
key_changes	-0.0045
note_above_threshold_set	-0.0116
percussive_presence	-0.0063
accented_Hzs_median	-0.0241

Jak widać na powyższym zestawieniu odłączenie każdej z cech powoduje nieznaczne zmiany dla wyniku modelu, co oznacza, że nie są one oderwana od rzeczywistości. Jednyna cecha która budzi wątpliwości to *repetitiveness* (powtarzalność), której usunięcie stanowiło poprawę *accuracy* modelu. Dalej będziemy rozważać czy taki wynik interpretować jako bodziec do pozbycia się tej cechy, czy może ustalimy, że lepiej ją zostawić.

Po przeprowadzeniu powyższego testu możnaby stwierdzić, że model bez cechy *repetitiveness* jest lepszy. Byłby to jednak pospieszony wniosek. Usunięcie po jednej z cech możemy rozumieć jako część działania brute force na całym zbiorze potęgowym cech, która to część obejmuje jedynie jednoelementowe podzbiory (równie dobrze moglibyśmy sprawdzać kombinacje np. 3 cech, których usunięcie daje najlepsze wyniki). Jako, że złożoność problemu obliczenia delt *accuracy* dla całego zbioru potęgowego nie wykracza poza sensowny zakres możliwości obliczeniowych ($card 2^X = 2^{12} = 4096$) sprawdzimy jaki dobór usuniętych cech da nam najlepsze wyniki:

dropped features sets	delta accuracy (pilot)	delta accuracy (cross-sectional)
tempo_variation key_changes note_above_threshold_set	0.0536	0.0021
tempo_variation median_spectral_rolloff_low_pitch key_changes note_above_threshold_set	0.0536	0.0009
tempo_variation repetitiveness median_spectral_rolloff_high_pitch key_changes note_above_threshold_set	0.0536	-0.0152
tempo_variation repetitiveness seconds_duration loudness_variation median_spectral_rolloff_high_pitch median_spectral_rolloff_low_pitch key_changes note_above_threshold_set	0.0536	-0.0596
tempo_variation note_above_threshold_set	0.0446	0.0042

Jak widać w dobranych pilotarzowo zestawach cech do usunięcia pojawiają się cechy występujące częściej (oznaczone kolorami). Wyniki poprawy *accuracy* dla pilotarzowej próbki przedstawiają niewielką poprawę (okolice 7%), które jednak przy zwiększeniu liczby próbek dla różnych ziaren praktycznie zeruje wynik poprawienia. Jednak jak widać są pewne cechy które powtarzają się często i można rozważyć ich usunięcie. Zmianę taką możnaby podjąć w celu chociażby uproszczenia modelu.

Dla następnych kroków będziemy brali jednak pod uwagę wszystkie *features*, ze względu na brak dostatecznej ilości przesłanek do wykluczenia żadnej z nich.

3.2.5 Badanie pobobieństw między gatunkami na podstawie modelu

Aby zbadać podobieństwo na bazie modelu obliczymy sumę pomyłek dla każdej pary kategorii. Jeśli suma równa się zero, nie zostanie pokazana na liście:

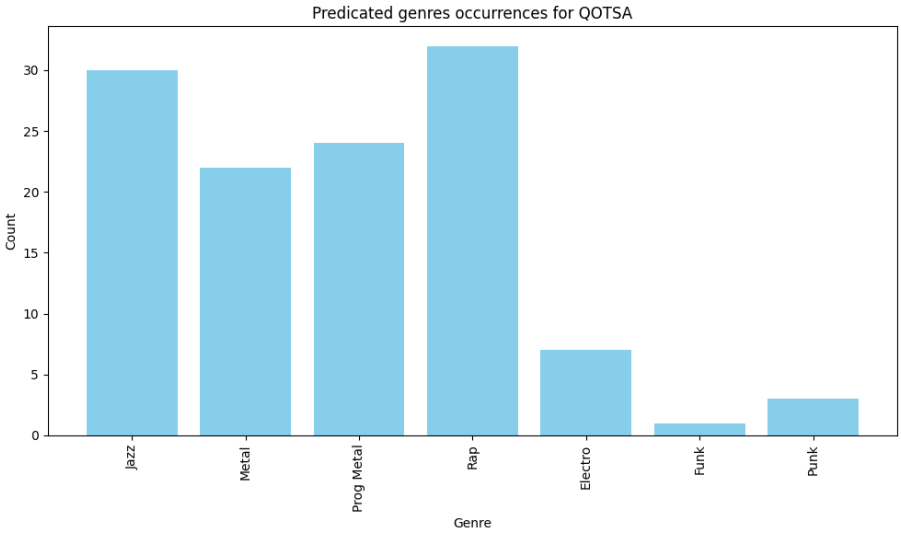
Genre Pairs	Count
Rap and Electro	4
Rap and Jazz	3
Jazz and Electro	2
Metal and Punk	2
Rap and Metal	2
Metal and Electro	1
Jazz and Funk	1
Prog Metal and Jazz	1
Prog Metal and Metal	1
Prog Metal and Rap	1
Electro and Punk	1

Model najwięcej razy pomylił Rap z Muzyką Elektroniczną i Rap z Jazzem. Jest to zarazem najciewsze powiązanie, bo daje nam wgląd w eklektyczną charakterystykę muzyki rapowej, która ewoluowała od *"Old School"*, który wywodził się właśnie z samplowanego Jazzu stopionego z uczuciami młodych artystów, żyjących w ciężkich warunkach, do "korporacyjnego" Rapu, napędzanego przez motywator zarobkowy, w którym widzimy obecnie mocne wpływy elektronicznej muzyki. Z innych ciekawych podobieństw mamy tu naprzykład Progresywny Metal i Jazz, które podobnie wykorzystują ciekawe i nietuzinkowe rozwiązania muzyczne.

3.2.6 Badanie klasyfikacji dyskografii *Queens of the Stone Age* na podstawie modelu

Badanie to jest szczególnie ciekawe przez nieoczywisty charakter muzyki *QOTSA*, spróbujemy wykorzystać model do określenia tego jak interpretuje gatunek poszczególnych piosenek, które bardzo różnią się odczuciem. W tej sekcji nie będziemy skupiać się na sprawdzaniu poprawności predykowanych gatunków, jedynie przeprowadzimy analizę rozkładu oraz przytoczymy skrajne przypadki ewidentnych pomyłek oraz przypadki ciekawe. To badanie ma na celu pozanie charakterystyki wspomnianej dyskografii przy pomocy narzędzia, którym będzie zaprojektowany wcześniej model (abstrahujemy od jego poprawności).

Po sklasyfikowaniu przez model rozkład poszczególnych klas dla dyskografii prezentuje się następująco:



Najczęstszym dopasowaniem dla dyskografii okazał się Rap, potem Jazz a następnie Metal Progresywny i metal. Zespół ten jest znany z balansowania między różnymi gatunkami, więc tak równy rozkład nie powinien dziwić. Ponadto również prawdą jest to, że w tej muzyce są wpływy funkowe, elektroniczne i punkowe, co również zostało uchwycone na wykresie. Muzyka klasyczna nie pojawiła się tu ani razu co jest zrozumiałe przez niewielką liczność kandydującej do tej klasy utworów.

Najciekawsze predykcje:

1. **I Think I Lost My Headache** — **Prog Metal**, w drugiej części piosenki występuje bardzo nowoczesny typ artystycznej "kakofonii", która może wyjaśniać tę predykcję.
2. **Little Sister** — **Metal**, przykładowa poprawna predykcja.
3. **Better Living Through Chemistry** — **Rap**, ten ciekawy utwór, odznaczający się swoją perkusyjną ścieżką mimo swojego mocno metalowego/alternatywnego charakteru został sklasyfikowany jako Rap, co pokazuje, jak powiązane są ze sobą te gatunki.
4. **Feet Don't Fail Me** — **Jazz**, ten utwór jest ciężki do klasyfikacji, ponieważ balansuje m.in. pomiędzy: Metalem, Bluesem, Funkiem i Electro. Ciekawym wyborem jest tutaj Jazz i subiektywny słuchacz może się zastanowić czy jest to sensowna predykcja, czy błąd modelu.
5. **This Lullaby** — **Electro**, jest to ewidentnie źle sklasyfikowany utwór. W rzeczywistości bardziej przypomina poezję śpiewaną.

4 Wnioski

Przebadane cechy gatunków muzycznych dały nam możliwość analizy problemów przedstawionych we wstępie 1. Z naszej analizy wynika, że wszystkie zaproponowane do ekstrakcji cechy wybroniły się 3.2.4 i nawet na niewielkim, napewno obarczonym zakłóceniami datasecie ich wypadkowa jest w stanie dość celnie (na poziomie **70% - 80%**) predykować gatunki muzyczne przy zastosowaniu wybranego modelu **"Gradient Boosting Classifier"** 3.2.3. Przeprowadzone badania ujawniły również ukryte podobieństwa między gatunkami 3.2.5 oraz pozwoliły na lepsze zrozumienie charakterystyki eklektycznej dyskografii zespołu *Queens of the Stone Age*. Należy pamiętać, że wszystkie wnioski zostały wyciągnięte na bazie modelu, który z definicji jest tylko przybliżeniem wycinka rzeczywistości i należy do wniosków podchodzić z pewną dozą dystansu.

Literatura

[1] Gradient boosting explanation. URL: `https://www.geeksforgeeks.org/ml-gradient-boosting/`.

[2] ABSounds. Musicgenreclassification. URL: `https://github.com/ABSounds/MusicGenreClassification`.

[3] egarciamartin. music-genre-classification-cnn. URL: `https://github.com/egarciamartin/music-genre-classification-cnn`.

A Dodatek

Kody źródłowe umieszczone zostały w repozytorium github:
`https://github.com/0leslaw/music_genres_analysis`.