

Politechnika Warszawska

WYDZIAŁ ELEKTRONIKI  
I TECHNIK INFORMACYJNYCH



Instytut Automatyki i Informatyki Stosowanej

# Praca dyplomowa magisterska

na kierunku Informatyka  
w specjalności Systemy Informacyjno-Decyzyjne

Badanie związku pomiędzy wydźwiękiem emocjonalnym tekstów  
piosenek a cechami statystycznymi melodii

**Mateusz Marzec**

Numer albumu 261692

promotor  
dr inż. Mariusz Kamola

WARSZAWA 2020



## Streszczenie

### **Badanie związku pomiędzy wydźwiękiem emocjonalnym tekstów piosenek a cechami statystycznymi melodii**

Praca magisterska opisuje zagadnienie klasyfikacji wydźwięku tekstów piosenek w języku angielskim oraz powiązanie jego wyników z cechami statystycznymi melodii. Podstawowym źródłem danych było *Million Song Dataset* oraz serwis *Spotify*. Klasyfikacja przeprowadzana jest z wykorzystaniem metody słownikowej oraz leksykonu *NRC*, a do jej ewaluacji wykorzystano zestaw etykiet zbudowany przy pomocy *tagów* z serwisu *last.fm*. Do przeprowadzenia badań skorzystano z cech statystycznych melodii udostępnianych przez serwis *Spotify*. Pierwsza część pracy koncentruje się na przedstawieniu metod analizy wydźwięku emocjonalnego tekstów oraz ich klasyfikacji. Następnie przedstawiono niezbędną teorię dotyczącą charakterystyki statystycznej melodii. Druga część prezentuje dokładny opis budowy projektu (klasyfikatora), wyniki oraz weryfikację jakości rozwiązania. Następnie uzyskane wyniki wykorzystano do przeprowadzenia analizy korelacji wydźwięku z cechami statystycznymi melodii. Na końcu pracy znajduje się krótkie podsumowanie, opisujące wyniki badań, co z nich wynika i jakie są dalsze możliwości analiz.

**Słowa kluczowe:** analiza wydźwięku, utwory muzyczne, leksykon, cechy statystyczne melodii

## **Abstract**

### **The study of correlation between sentiment of song lyrics and statistical features of the melody**

The master's thesis describes issues of mood classification based on English song lyrics and correlation of its results with statistical features of the melody. Study is based on *Million Song Dataset* and *Spotify API*. Classification is carried out using the dictionary method and the *NRC* lexicon. Classifier was evaluated by a set of labels built using *tags* from the *last.fm* website. The statistical features of the melody was provided by the *Spotify* app. The first part of document focuses on presenting methods for sentiment analysis and classification of English text. Following part of thesis describes theory of statistical characteristics of the melody. Next part presents a details about main project (classifier), results and evaluation of classification quality. Afterwards, the obtained results were used to analyze the correlation of lyrics sentiment with statistical features of the melody. Thesis ends with short summary describing the results of the study and further possibilities of development.

**Keywords:** sentiment analysis, songs, lexicon, statistical features of the melody



**Politechnika Warszawska**

załącznik nr 3 do zarządzenia  
nr 28 /2016 Rektora PW

.....  
miejsowość i data

.....  
imię i nazwisko studenta

.....  
numer albumu

.....  
kierunek studiów

### **OŚWIADCZENIE**

Świadomy/-a odpowiedzialności karnej za składanie fałszywych zeznań oświadczam, że niniejsza praca dyplomowa została napisana przeze mnie samodzielnie, pod opieką kierującego pracą dyplomową.

Jednocześnie oświadczam, że:

- niniejsza praca dyplomowa nie narusza praw autorskich w rozumieniu ustawy z dnia 4 lutego 1994 roku o prawie autorskim i prawach pokrewnych (Dz.U. z 2006 r. Nr 90, poz. 631 z późn. zm.) oraz dóbr osobistych chronionych prawem cywilnym,
- niniejsza praca dyplomowa nie zawiera danych i informacji, które uzyskałem/-am w sposób niedozwolony,
- niniejsza praca dyplomowa nie była wcześniej podstawą żadnej innej urzędowej procedury związanej z nadawaniem dyplomów lub tytułów zawodowych,
- wszystkie informacje umieszczone w niniejszej pracy, uzyskane ze źródeł pisanych i elektronicznych, zostały udokumentowane w wykazie literatury odpowiednimi odnośnikami,
- znam regulacje prawne Politechniki Warszawskiej w sprawie zarządzania prawami autorskimi i prawami pokrewnymi, prawami własności przemysłowej oraz zasadami komercjalizacji.

Oświadczam, że treść pracy dyplomowej w wersji drukowanej, treść pracy dyplomowej zawartej na nośniku elektronicznym (płycie kompaktowej) oraz treść pracy dyplomowej w module APD systemu USOS są identyczne.

.....  
czytelny podpis studenta



# Spis treści

<b>1. Wstęp</b>	9
<b>2. Cel i zakres pracy</b>	10
<b>3. Analiza wydźwięku tekstów</b>	11
3.1. Klasyfikacja wydźwięku	11
3.1.1. Klasyfikacja binarna	11
3.1.2. Klasyfikacja ciągła	12
3.1.3. Klasyfikacja przestrzenna	12
3.2. Wyzwania analizy wydźwięku	13
3.2.1. Złożoność i różnorodność języków	13
3.2.2. Negacje	13
3.2.3. Ironie	14
3.2.4. Błędy ludzkie	14
3.3. Metody analizy wydźwięku	14
3.3.1. Metoda ręczna	14
3.3.2. Metody słownikowe	15
3.3.3. Metody statystyczne	16
3.4. Zastosowanie analizy wydźwięku	16
<b>4. Klasyfikacja tekstów piosenek</b>	18
4.1. Pozyskanie danych	18
4.1.1. Teksty utworów	18
4.1.2. Selekcja danych	18
4.1.3. Etykietyzacja	18
4.2. Obróbka danych	19
4.2.1. Balansowanie danych	19
4.2.2. Tokenizacja	20
4.2.3. Filtracja <i>stop words</i> oraz ujednocnianie słów	20
4.3. Ekstrakcja cech	21
4.3.1. Word embedding	21
4.3.2. Ważenie częstości termów	22
4.3.3. Bag of words (BOW)	23
4.4. Analiza wydźwięku	24
4.4.1. Wybór klasyfikatora	24
4.4.2. Budowa modelu	25
4.5. Ewaluacja modelu	25
4.5.1. Dokładność	26
4.5.2. Macierz błędów	26
<b>5. Analiza statystyczna melodii</b>	27
5.1. Cechy melodii	27
5.1.1. Rytm	27

## 0. Spis treści

---

5.1.2.	Dynamika . . . . .	28
5.1.3.	Charakterystyka tonalna . . . . .	28
5.1.4.	Barwa . . . . .	28
5.2.	Źródła charakterystyki statystycznej muzyki . . . . .	28
5.3.	Klasyfikacja z wykorzystaniem cech statystycznych . . . . .	30
<b>6.</b>	<b>Projekt . . . . .</b>	<b>31</b>
6.1.	Źródła danych . . . . .	31
6.1.1.	Million Song Dataset . . . . .	31
6.1.2.	Lyric Wikia . . . . .	32
6.1.3.	Spotify API . . . . .	32
6.1.4.	Leksykon <i>NRC</i> . . . . .	33
6.1.5.	Tagi <i>last.fm</i> . . . . .	36
6.2.	Analiza przy użyciu metody słownikowej . . . . .	36
6.2.1.	Obróbka danych . . . . .	37
6.2.2.	Ekstrakcja cech . . . . .	39
6.2.3.	Klasyfikacja . . . . .	40
6.2.4.	Wyniki . . . . .	41
6.2.5.	Ewaluacja rozwiązania . . . . .	41
6.3.	Analiza związku cech statystycznych melodii z wynikami klasyfikacji . . . . .	43
6.3.1.	Badanie korelacji . . . . .	43
6.3.2.	Skorelowane cechy statystyczne a wydźwięk utworów . . . . .	45
6.3.3.	Weryfikacja nieskorelowanej cechy . . . . .	49
6.4.	Podsumowanie wyników . . . . .	51
6.5.	Wykorzystane narzędzia oraz współczynniki . . . . .	52
6.5.1.	NLTK . . . . .	52
6.5.2.	Sklearn . . . . .	52
6.5.3.	PyLyrics . . . . .	52
6.5.4.	SciPy . . . . .	53
6.5.5.	Współczynnik Pearsona . . . . .	53
6.5.6.	Współczynnik Spearmana . . . . .	53
<b>7.</b>	<b>Zakończenie . . . . .</b>	<b>54</b>
	<b>Bibliografia . . . . .</b>	<b>56</b>
	<b>Spis rysunków . . . . .</b>	<b>59</b>
	<b>Spis tabel . . . . .</b>	<b>60</b>



# 1. Wstęp

Muzyka jest jedną z najpopularniejszych dziedzin sztuki. Towarzyszy ludzkości od tysiącleci, początkowo wykorzystywana dla celów religijnych [1]. Za najstarszy, wiernie odtworzony współcześnie, utwór muzyczny uważany jest *Hymn do Nikkal* pochodzący z 1400 roku p.n.e. Obecnie muzyka spełnia głównie cele rozrywkowe. Jej popularność sprawiła, że stała się jednym z głównych sposobów na przekazywanie uczyć oraz wyrażanie opinii na określone tematy. W związku z tym współczesne utwory charakteryzują się wysokim nacechowaniem emocjonalnym [2].

Ze względu na środki wykonawcze muzykę możemy podzielić na trzy grupy: wokalne, instrumentalne oraz wokально-instrumentalne, z czego ostatnia z nich jest współcześnie najpopularniejsza.

W utworach wolno-instrumentalnych głównym nośnikiem emocji są słowa [2]. Ich treść w znacznym stopniu wpływa na odbiór piosenki. Nie bez znaczenia jest, także melodia im towarzysząca. Żywa, dynamiczna linia melodyczna napędza energią, natomiast spokojna i cicha działa uspokajająco. Autorzy wykorzystują określone cechy statystyczne melodii aby pogłębić odczucia, wynikające z tekstów lub wprowadzić słuchaczy w konsternację. Z pośród nich wyróżnić można rytm, tempo, głośność czy energię, które najbardziej wpływają na odbiór melodii [3].

Celem przedstawionej pracy jest zbadanie zależności między wydźwiękiem emocjonalnym tekstów utworów muzycznych a cechami statystycznymi melodii. Przedstawiona analiza koncentruje się na piosenkach anglojęzycznych - ze względu na dużą dostępność danych - publicznie dostępnych w serwisie *Million Song Dataset*<sup>1</sup>.

Jednoznaczne określenie wydźwięku emocjonalnego tekstu utworu nie jest jednak proste. Zawierają one wiele słów nacechowanych emocjonalnie, często sprzecznie. Dodatkowo, ze względu na prawa autorskie, nie są one tak łatwo dostępne. Szczegółowy opis wyzwań z jakimi mierzyć się musi analiza wydźwięku oraz sposoby ich rozwiązywania przedstawiona została w dalszej części pracy.

Analiza wydźwięku, a co za tym idzie, klasyfikacja utworów muzycznych jest tematem wielu badań oraz publikacji. Można je podzielić - z punktu widzenia implementacyjnego - na bazujące na tekstach utworów oraz cechach statystycznych melodii. Pierwsze z nich wymagają skomplikowanego procesu przygotowania słów oraz ekstrakcji cech, aby można było je wykorzystać np. podczas budowy modelu uczenia maszynowego [4]. Drugie, z uwagi na liczbowy charakter danych, nie wymagają wcześniejszej obróbki [5]. Dostępne są także badania porównujące obydwa podejścia, wyróżniające ich zalety oraz wady [6].

---

<sup>1</sup> <http://millionsongdataset.com/>

## 2. Cel i zakres pracy

Celem pracy jest zbadanie stopnia zależności wydźwięku emocjonalnego tekstów współczesnych utworów muzycznych od cech statystycznych melodii. W skład pracy wchodzi również analiza wpływu charakterystyki melodycznej na ogólny odbiór emocjonalny piosenek. Ponadto przedstawiono zestawienie metod analizy wydźwięku emocjonalnego tekstów. Praca bazuje na piosenkach anglojęzycznych, publicznie dostępnych w serwisie *Million Song Dataset* oraz cechach statystycznych udostępnionych przez aplikację *Spotify*<sup>2</sup>.

Pierwszy rozdział jest wstępem wprowadzającym w tematykę muzyki oraz jej wpływu na emocje słuchacza.

Rozdział drugi przedstawia cel i zakres pracy, wskazuje główne źródła danych jakie wykorzystywane będą w projekcie magisterskim.

Trzeci rozdział opisuje w szerszym zakresie problematykę analizy sentymentu tekstów. Wskazuje popularne metody analizy wydźwięku, przedstawia jej współczesne wyzwania oraz zastosowania.

Czwarty rozdział jest przedstawia przykładowy proces klasyfikacji tekstu utworu, z zastosowaniem nadzorowanego uczenia maszynowego, do wcześniej zdefiniowanych klas. Opisuje każdy z etapów, począwszy od zebraniu danych, na właściwej analizie skończywszy. Dodatkowo sekcja prezentuje różne podejścia do ekstrakcji cech oraz najpopularniejsze klasyfikatory.

Piąty rozdział poświęcony jest analizie statystycznej melodii. Wskazuje jej popularne cechy, sposób ich pozyskiwania oraz przedstawia metodykę klasyfikacji utworów z ich wykorzystaniem.

Rozdział szósty opisuje dokładnie projekt magisterski. Prezentuje sposób pozyskiwania oraz obróbki danych. Przedstawia metodykę oraz wyniki analizy, wykorzystane narzędzia i szczegóły implementacyjne.

Siądmy rozdział stanowi zakończenie pracy, przedstawia wnioski wyciągnięte z uzyskanych wyników, podsumowuje wykonaną pracę i wskazuje możliwe dalsze etapy badań.

---

<sup>2</sup> <http://spotify.com/>

### 3. Analiza wydźwięku tekstów

Tekst jest nośnikiem wielu informacji, co sprawia, że stał się obiektem wielu rodzajów analiz. Od pozyskania prostych informacji tekstowych (*Information Extraction*), przez odkrywanie modeli koncepcyjnych i wiedzy zawartej w tekstach (*Knowledge Discovery in Databases*), po klasyfikację (*Text Classification*) [7]. Jednym ze sposobów klasyfikacji tekstów jest podział ze względu na wydźwięk emocjonalny, wykorzystuje się w tym celu analizę wydźwięku.

Podstawowymi zadaniami analizy wydźwięku jest identyfikacja oraz klasyfikacja fragmentów lub całych wypowiedzi [7]. Koncentruje się na wydobyciu wartości emocjonalnej z danego tekstu. Analiza wydźwięku opiera się na dwóch założeniach. Po pierwsze zakłada, że niektóre słowa, zawarte w tekście, wyrażają emocje, po drugie, wypowiedzenie części z nich wywołuje emocje. Badaniu poddawane są więc zarówno terminy świadczące o stanie emocjonalnym autora jak i te mające wywołać emocje u odbiorcy. Z uwagi na subiektywny charakter oraz złożoność badanych danych, określanie wydźwięku jest procesem skomplikowanym i często wymagającym dużej ilości próbek.

#### 3.1. Klasyfikacja wydźwięku

Wynikiem analizy wydźwięku jest m.in. klasyfikacja fragmentu tekstu lub całości do jednej z kategorii zależnej od przyjętego modelu [8]. Analiza wydźwięku najczęściej sprowadza się do określenia czy dana treść jest pozytywna czy też negatywna [9]. Występuje jednak wiele innych reprezentacji, charakteryzujących się różnymi zaletami oraz wadami. Bardziej szczegółowo zostały one przedstawione poniżej.

##### 3.1.1. Klasyfikacja binarna

Tekst klasyfikowany jest do jednej z dwóch kategorii: pozytywny lub negatywny. Jest to najpopularniejsze rozwiązanie ze względu na swoją prostotę. Bardzo dobrze pasuje do analiz krótkich tekstów np. opinii. Reprezentacja ta jednak nie niesie informacji o wadze wydźwięku, jest mało precyzyjna i często niewystarczająca do niektórych zastosowań analitycznych. Dla przykładu:

*To dobry produkt*

oraz

*To **bardzo** dobry produkt*

zostaną zakwalifikowane do tej samej kategorii (*wydźwięk pozytywny*). Powyższa reprezentacja nie wskaże żadnej różnicy w wydźwięku pomimo jej występowania.

#### 3.1.2. Klasyfikacja ciągła

Przedstawienie wydźwięku emocjonalnego tekstu w postaci ciągłej skali ocen np. 5/10. Pozwala na szczegółowe określenie jego wagi, jednakże uniemożliwia grupowanie. Wykorzystywana jest często w leksykonach zawierających wartości emocjonalne terminów.

Dla przykładu z poprzedniego punktu, reprezentacja ciągła, w przeciwieństwie do binarnej, wskaże różnicę w wydźwięku np. kolejno 6/10 oraz 8/10.

Oprócz wspomnianego wcześniej problemu z grupowaniem, tego typu klasyfikacja nie uszczegóławia wydźwięku. Przykładowo niezależnie czy dany termin może świadczyć o smutku czy złości, zakwalifikowany zostanie jako wartość negatywna.

#### 3.1.3. Klasyfikacja przestrzenna

Wykorzystanie przestrzeni dwu lub więcej wymiarowej do określenia wydźwięku. Każdy z wymiarów odzwierciedla natężenie wybranego parametru, a wydźwiękiem jest punkt w tejże przestrzeni [10]. Przykładem takiego podejścia jest *model Russella*, wykorzystujący dwuwymiarową przestrzeń opisywaną przez *walencyjność* oraz *pobudzenie*. Pozwala on na klasyfikację tekstów np. na cztery grupy, z których każda odpowiada jednej ćwiartce przestrzeni *Russella*. Na rysunku 3.1 przedstawiono wizualizację przestrzeni walencyjno-pobudzeniowej.



**Rysunek 3.1.** Wizualizacja przestrzeni walencyjno-pobudzeniowej zaproponowanej przez *Russella*

Takie podejście umożliwia znacznie bardziej szczegółową reprezentację wydźwięku emo-

cjonalnego, niż binarna i jest wykorzystywane np. do podziału utworów muzycznych ze względu na nastrój.

## 3.2. Wyzwania analizy wydźwięku

Tak jak wcześniej wspomniano, badanie wydźwięku emocjonalnego tekstu niesie ze sobą wiele problemów, wyzwań których rozwiązanie nie jest trywialne i od dłuższego czasu jest przedmiotem wielu publikacji [9][2]. Najczęściej spotykane bariery, przedstawiono poniżej.

### 3.2.1. Złożoność i różnorodność języków

Jest to jedno z największych wyzwań, w szczególności dla analiz wydźwięku tekstów nieanglojęzycznych. Języki różnią się ze względu na słownictwo, gramatykę oraz alfabet co implikuje konieczność tworzenia nowych słowników, zbiorów danych uczących, reguł itp. specyficznych dla każdego z nich. Należy też pamiętać o tym, że języki podlegają ewolucji, zmianom pokoleniowym, które wymuszają ich ciągłą aktualizację.

### 3.2.2. Negacje

Niezwykle istotnym wyzwaniem, w szczególności dla analizy krótkich tekstów (np. opinii), jest detekcja negacji, która może całkowicie zmienić wydźwięk badanych treści [11]. Dla przykładu:

*Ten telewizor **nie** jest **zły***

Nie biorąc pod uwagę negacji, powyższa opinia zostanie zakwalifikowana jako negatywna, co jest oczywistym błędem. Rozwiązaniem tego problemu jest wyszukiwanie słów negujących przed terminami o nacechowaniu emocjonalnym. W przypadku ich znalezienia, polaryzacja wydźwięku jest odwracana, dzięki czemu uzyskany zostaje prawidłowy sens wypowiedzi. Takie podejście niesie ze sobą kolejne wyzwania, ponieważ negacja nie zawsze znajduje się dokładnie za negowanym słowem (w praktyce zdarza się to bardzo rzadko), dodatkowo czasem odnosi się do jednego a czasem do wielu przymiotników. Przykładowo:

*Ten telewizor **nie** jest **zły** i **drogi***

Słowo negujące *nie* odnosi się w tym wypadku do obydwu przymiotników ale nie zawsze tak musi być.

To jak daleko należy szukać negacji od słowa nacechowanego emocjonalnie i jak ją dokładnie interpretować jest przedmiotem wielu badań co świadczy o złożoności problemu.

#### 3.2.3. Ironie

Detekcja wypowiedzi ironicznych jest niezwykle trudnym wyzwaniem, często nawet dla ręcznej analizy wydźwięku. Określenie prawdziwych intencji autora jest kluczowe dla określenia sensu wypowiedzi [12]. Dla przykładu:

*Ten telewizor jest **wspaniały!** Jest zdecydowanie zbyt **tani***

Na pierwszy rzut oka, powyższa wypowiedź wydaje się być pozytywna, wynika to z faktu wystąpienia słów o wydźwięku pozytywnym. Jednakże analizując cały kontekst wypowiedzi widać, że prawdziwy sens jest zgoła inny.

To co dodatkowo utrudnia powyższe zagadnienie, to występowanie wielu typów sarkazmu, powiązanych z językiem oraz kulturą autora. Widać więc, że ironie są niezwykle trudnym problemem, mogącym w znacznym stopniu wpłynąć na ogólny wydźwięk tekstu.

#### 3.2.4. Błędy ludzkie

Dane wejściowe do analizy wydźwięku nie są pozbawione błędów, głównie ortograficznych, gramatycznych oraz interpunkcyjnych. Pierwsze w nich bywają kłopotliwe, jeżeli pomyłka uniemożliwia jednoznaczną identyfikację słowa. Pozostałe znacznie pogarszają jakość analizy dla metod bazujących na regułach gramatycznych i interpunkcyjnych.

### 3.3. Metody analizy wydźwięku

Analiza wydźwięku ma wiele różnych implementacji, które różnią się m.in. poziomem skomplikowania, trafnością czy też czasochłonnością. Poniżej przedstawiono szereg najpopularniejszych z nich, pogrupowanych ze względu na podejście do problemu, z podkreśleniem zalet oraz wad [13] [7].

#### 3.3.1. Metoda ręczna

Najprostsza i najbardziej prymitywna metoda. Nie wymaga implementacji, natomiast wymaga zaangażowania jednostek ludzkich odpowiedzialnych za jej przeprowadzenie. Ich zadaniem jest subiektywna ocena emocji jakie niesie ze sobą utwór. Jej największą zaletą jest brak konieczności zastosowania skomplikowanych narzędzi do automatycznego określania wydźwięku i bardzo dobra jakość uzyskanych wyników, jednakże z punktu widzenia analizy dużych zbiorów danych jest praktycznie bezużyteczna.

### 3.3.2. Metody słownikowe

Metody te koncentrują się na analizie zawartości tekstów i wypowiedzi. Wykorzystują słowa kluczowe, biorą pod uwagę reguły leksykalne i składniowe używanego języka [14]. W zależności od wybranej metody, niezbędna jest znajomość gramatyki, znaczenia słów lub specyfiki wypowiedzi.

Metoda wykorzystująca słowa i wyrażenia kluczowe bazuje na słownikach wydźwięków (leksykonach), które każdemu terminowi, w sposób jednoznaczny lub probabilistyczny przypisują określoną wartość emocjonalną. Typ klasyfikacji przeprowadzonej przy pomocy takiej metody zależy od wykorzystanego słownika. Klasyfikacja przeprowadzana jest poprzez wyszukiwanie w treści słów o zabarwieniu emocjonalnym, sprawdzeniu ich wydźwięku w leksykonie i na tej podstawie klasyfikacja tekstu. W sytuacji gdy odnaleziono zostanie więcej niż jeden termin, wynikiem analizy jest ich uśredniona wartość. Dla przykładu:

*To **dobry** produkt, **szkoda** że taki **drogi***

Powyższy tekst, dla klasyfikacji binarnej, zostanie opisany jako negatywny. Wynika to z przewagi terminów o takim też wydźwięku.

Warto zauważyć, że w przypadku tej metody, problemem może być niejednoznaczny wydźwięk niektórych słów. Przeciwdziała się temu poprzez wykorzystanie leksykonów zawierających prawdopodobieństwa danego nacechowania emocjonalnego.

Zaletą powyższej metody jest prostota implementacji, powszechna dostępność leksykonów (w szczególności w języku angielskim) oraz mnogość gotowych rozwiązań. Wadą natomiast jest problem z klasyfikacją ironii, negacji oraz niejednoznaczność niektórych słów kluczowych.

Metoda wykorzystująca reguły leksykalne jest bardziej zaawansowana niż poprzednia. Analizuje tekst jako całość [7]. Klasyfikuje treści przy pomocy wzorców odnoszących się do następujących obszarów:

- gramatyka języka - miejsce słów w zdaniu
- podobieństwa i różnice w znaczeniu słów - synonimia, antonimia, hiperonimia, homonimia, meronimia
- zasób słów danego języka

Tego typu rozwiązania nie są jednak uniwersalne, zestawy reguł często trzeba definiować specjalnie do badanego problemu, wymaga duże wiedzy językowej i lingwistycznej, która nie jest trywialna.

#### 3.3.3. Metody statystyczne

Tego typu podejście wykorzystuje techniki ilościowe, traktujące tekst jako obiekt charakteryzowany za pomocą danych ilościowych [15][7]. Dane te opisywać mogą zarówno zawartość dokumentu (liczba słów, fraz, poziom złożoności itp.) jak i sam dokument (liczba znaków w dokumencie, data publikacji itp.). W praktyce, badany tekst reprezentowany jest jako wektor w wielowymiarowej przestrzeni, opisywanej przez zbiór cech.

Metody statystyczne wykorzystujące uczenie maszynowe stają się coraz bardziej popularne. Powodem jest ich uniwersalność i szerokiej dostępności narzędzi. W ogólności, klasyfikator, zbudowany na bazie uczenia maszynowego, ma za zadanie nadać odpowiednią etykietę (zakwalifikować) elementom wejściowym. W przypadku analizy wydźwięku elementem wejściowym będzie tekst, natomiast etykietą wydźwięk emocjonalny.

W ramach klasyfikacji przy pomocy metod statystycznych wyróżnić można dwa podstawowe podejścia wykorzystujące:

- uczenie nadzorowane (z nauczycielem)
- uczenie nienadzorowane (bez nauczyciela)

Pierwsze z nich bazuje na klasyfikatorze uczonym przy pomocy specjalnie przygotowanych danych treningowych, w skład których wchodzi teksty oraz etykiety w postaci wydźwięku. Przykładem tego typu klasyfikatorów są drzewa i reguły decyzyjne, sieci neuronowe, metody bayesowskie i metody regresywne. Problemem tego podejścia jest etykietyzacja danych, wymaganych do stworzenia odpowiednio dużego zbioru danych trenujących i testujących.

Drugie podejście wymaga od klasyfikatora autonomicznego wykrycia wzorców oraz reguł w zbiorze danych. Przykładowe techniki stosowane w nim to algorytm centroidów, metody hierarchiczne, oparte na miarach podobieństwa i miarach złożoności tekstów.

#### 3.4. Zastosowanie analizy wydźwięku

Teksty nacechowane emocjonalnie są ogólnodostępne np. w postaci tweetów, opinii, tekstów piosenek, artykułów prasowych itp., pozyskanie ich w dużej ilości nie sprawia problemu co wprost implikuje szerokie zastosowanie analizy wydźwięku zarówno w badaniach tak i w wielu dziedzinach życia [16]. Poniżej przedstawiono najpopularniejsze z nich.

- Analiza opinii klientów np. sklepów internetowych. Pozwala na tanią i precyzyjną informację na temat ich zdania, bez wykorzystania drogich zewnętrznych firm analitycznych lub ankiet.



- Detekcja nienawiści, zachowań rasistowskich i ksenofobicznych w internecie. Bardzo popularne zastosowanie w związku z wszechobecnym hatem.
- Grupowanie utworów pod kątem nastroju. Często wykorzystywane przez wszelkiego rodzaju serwisy streamingowe, pozwala na podpowiadanie użytkownikowi piosenek w zależności od wybranego nastroju.
- Badanie nastrojów społecznych przy pomocy postów na serwisach społecznościowych. Szczególnie często wykorzystywane przez partie i organizacje polityczne
- Systemy rekomendacyjne, pozwalające na proponowanie internatom produktów w zależności od wystawionych przez nich opinii.
- Filtrowanie spamu oraz maili o podejrzanym treści.

## 4. Klasyfikacja tekstów piosenek

W tym rozdziale przedstawiony zostanie proces klasyfikacji utworów muzycznych ze względu na wydźwięk emocjonalny ich tekstów. Opisane zostaną jej kolejne etapy od wyselekcjonowania danych, przez ich przygotowanie, ekstrakcję cech aż do docelowej analizy i interpretacji wyników [6].

Przedstawiona klasyfikacja wydźwięku tekstów piosenek bazuje na nadzorowanym uczeniu maszynowym, przedstawionych w poprzednim rozdziale. Utwory przydzielane są do jednej z czterech kategorii zaproponowanych w modelu *Russella* [10]. Cechą charakterystyczną tego typu badań jest długość danych wejściowych - znakomita większość utworów posiada długość od kilkudziesięciu do kilkuset słów, w związku z czym ogólny wydźwięk emocjonalny jest złożeniem wielu składowych.

### 4.1. Pozyskanie danych

Klasyfikacja z wykorzystaniem nadzorowanego uczenia maszynowego wymaga odpowiedniego zbioru danych treningowych, testowych oraz walidacyjnych, którym w tym wypadku będą teksty utworów muzycznych opatrzone etykietami określającymi wzorcowy wydźwięk emocjonalny.

#### 4.1.1. Teksty utworów

Najczęściej wykorzystywanym źródłem w klasyfikatorach tekstów piosenek jest *Million Song Dataset* darmowy zbiór posiadający około miliona rekordów, z których każdy zawiera podstawowe informacje o utworze - nazwę, autora, rok produkcji, gatunek, wytwórnę, nazwę albumu itp. [4]. Wskazane baza nie zawiera tekstów piosenek z uwagi na prawa autorskie. W celu ich uzyskania wykorzystuje się zewnętrzne biblioteki np. *PyLyrics*.

#### 4.1.2. Selekcja danych

Od wspomnianego wcześniej zbioru danych należy odfiltrować wszystkie utwory nieposiadające tekstu oraz - ze względu na ujednolicenie języka - te których tekst jest w innym języku niż angielski. Wykorzystuje się w tym celu biblioteki posiadające wbudowane słowniki. Przykładem takiego słownika, dla języka angielskiego, jest *corpus* biblioteki *NLTK*.

#### 4.1.3. Etykietyzacja

Znaczącym wyzwaniem jest uzyskanie etykiet. Nie występuje ogólnie dostępny oraz wystarczająco duży zbiór etykiet - określających wydźwięk emocjonalny - przypisanych

do konkretnych utworów. Aby rozwiązać ten problem najczęściej wykorzystuje się serwisy muzyczne np. *last.fm*, pozwalające użytkownikom na określenie emocji, które towarzyszą im podczas słuchaniu danego utworu [17]. Tego typu źródło etykiet niesie ze sobą pewne zagrożenie. Przy określeniu emocji słuchacze często kierują się nie tylko tekstem danego utworu, ale także melodią, etykiety odnoszą się więc do wydźwięku całego utworu a nie jedynie tekstu.

Etykiety tworzy się analizując *tagi* pozostawione przez użytkowników portalu. Z pośród nich wyszukuje się te, które określają emocje, następnie grupuje się do zdefiniowanych wcześniej kategorii [17]. Przykładowa lista tagów uzyskana z serwisu *last.fm*:

calm, comfort, quiet, happy, happiness, sad, sadness, aggression

Powyższe tagi pogrupować można np. na cztery kategorie:

relaks, szczęście, smutek, złość

W ten sposób można określić dominujący wydźwięk w utworze, który staje się jego etykietą.

W przypadku gdy utwór nie posiada tagów charakteryzujących emocje, jest usuwany ze zbioru danych. W praktyce występuje to dość często z uwagi na bardzo dużą ilość rekordów i stosunkowo niewielką liczbę tagów użytkowników.

## 4.2. Obróbka danych

Po uzyskaniu wymaganego zbioru danych wykonuje się serię operacji, aby przygotować je do późniejszej analizy [18]. Poniżej przedstawiono najczęściej wykorzystywane techniki wstępnej obróbki tekstów piosenek.

### 4.2.1. Balansowanie danych

Aby zapewnić jak najwyższą jakość klasyfikacji, należy doprowadzić do zbalansowania dystrybucji wydźwięku emocjonalnego w danych [19]. Dla przykładu mając rozkład uzyskanych etykiet w zbiorze danych przedstawiony w tabeli 4.1 balans nie jest zachowany.

Relaks	Szczęście	Smutek	Złość
3000	2000	1000	1500

**Tabela 4.1.** Rozkład etykiet emocjonalnych w przykładowym zbiorze danych

Występuje znacznie więcej utworów opisanych przez etykiety jako *relaksujące* niż pozostałych. Balansowanie wykonuje się poprzez kopiowanie rekordów dla deficytowych etykiet i doklejanie ich do istniejącego zbioru tak aby dla każdej klasy wydźwięku liczba

piosenek była ta sama [19]. Dodatkowo tekst w skopiowanych utworach każdorazowo zostaje wymieszany, aby nie występowały identyczne rekordy.

##### 4.2.2. Tokenizacja

Tokenizacja jest procesem polegającym na podziale monolitycznego tekstu na zadane części składowe - *tokens* [20]. Najczęściej pojedynczym *tokenem* jest słowo. Przykładem implementacji funkcji tokenizującej jest *tokenize* biblioteki *NLTK*.

##### 4.2.3. Filtracja *stop words* oraz ujednolicanie słów

Część słów zawartych w tekstach utworów nie posiada wartości z punktu widzenia analizy wydźwięku (tzw. *stop words*). Jeżeli metoda analizy nie wykorzystuje składni i struktury zdania - jak w tym wypadku - można usunąć je z danych. W tym celu korzysta się ze ogólnodostępnych list *stop words* dla interesującego nas języka [21].

Kolejnym etapem jest ujednolicanie słów [22]. Wykorzystuje się w tym celu jedną z dwóch operacji:

- *stemming* - wydobycie z wyrazu *stema* - rdzenia, który nie podlega odmianom
- lematyzacja - sprowadzenie słowa do jego formy podstawowej - leksemu

W tabeli 4.2 przedstawiono przykłady wyników wykorzystania obydwu metod.

Słowo	Stem	Leksem
wytrwale	wytrwał	wytrwały
roweru	rower	rower
wiórkami	wiór	wiórek
robienie	robi	robi

**Tabela 4.2.** Lista słów przed i po przeprowadzeniu *stemmingu* oraz lematyzacji

Powyższe operacje znacząco zmniejszą liczbę słów w słowniku, co zwiększa ich reprezentacyjność w korpusie.

Filtrowanie *stop words* oraz ujednolicenie nie są obowiązkowe jednakże pozwalają w znaczący sposób prawić jakoś klasyfikacji [19].

Warto zauważyć, że powyższe operacje przygotowywania danych są dość uniwersalne, wykonywane nie tylko w analizie wykorzystującej uczenie maszynowe ale także metody słownikowe.

### 4.3. Ekstrakcja cech

Po uzyskaniu wymaganego zbioru danych oraz poddaniu go wstępnej obróbce należy przekształcić go do formy zrozumiałej dla klasyfikatora - wektora cech. W tym celu wykonuje się tzw. ekstrakcje cech. Proces ten wykonać można na wiele sposobów w zależności od tego jakie cechy z tekstu chcemy uzyskać [23]. Poniżej przedstawiono najpopularniejsze, stosowane podczas klasyfikacji tekstów piosenek.

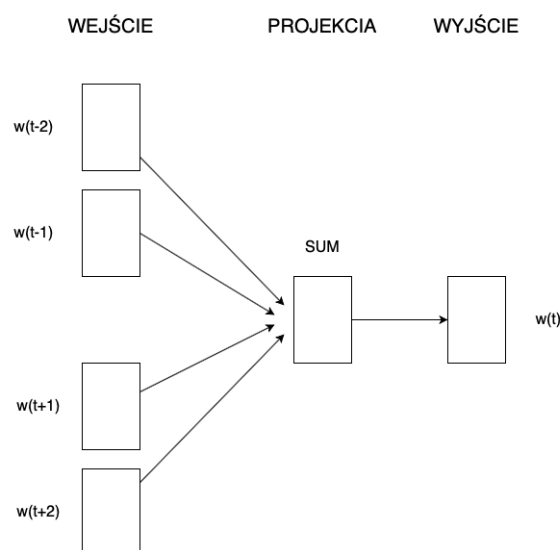
#### 4.3.1. Word embedding

*Word embedding* jest sposobem reprezentacji wyrazów w postaci liczbowej, potrzebnej do zastosowania uczenia maszynowego. Obecnie najczęściej wykorzystuje się dwuwarstwową sieć neuronową *word2vec*, która na wejściu przyjmuje korpus tekstowy, a na wyjściu dostarcza zbiór wektorów cech, reprezentujący słowa z tegoż korpusu i opisujących powiązania między nimi [24].

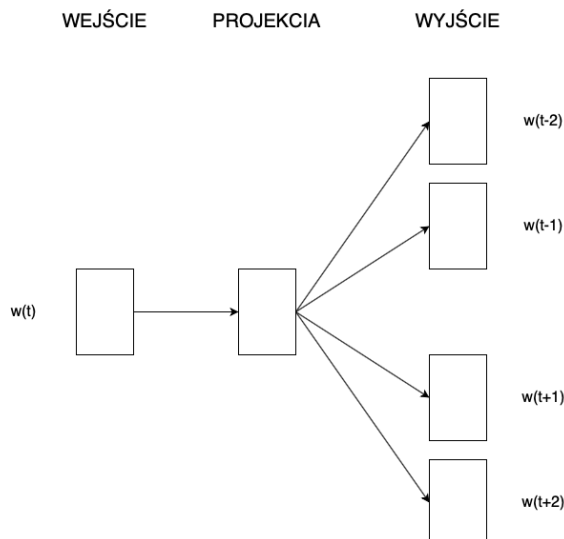
Word2vec może wykorzystywać jedną z dwóch architektur, metod do stworzenia reprezentacji słów:

- Continuous bag of words (CBOW) - model przewiduje aktualnie uczone słowo na podstawie słów otaczających bez uwzględnienia kolejności (bag of words)
- Continuous skip-gram - model wykorzystuje aktualne słowo do prognozowania słów otaczających, przy czym słowom bliższym przyporządkowana jest wyższa waga

Podejście *CBOW* zaprezentowane jest na rysunku 4.1 natomiast *Continuous skip-gram* na 4.2.



Rysunek 4.1. Wizualizacja modelu *CBOW* w *word2vec*



**Rysunek 4.2.** Wizualizacja modelu *Continuous skip-gram* w *word2vec*

Pierwsza z nich cechuje się większą szybkością działania, natomiast druga lepszą precyzją, w szczególności dla mniej popularnych słów.

Podczas tworzenia *word embedding*'u należy wybrać wymiarowość przestrzeni (zwykle jest to 100 lub 300) oraz rozmiar okna w jakim *word2vec* będzie próbował znaleźć słowa. Przykładową implementację *word2vec* w języku *Python* oferuje biblioteka *Gensim*.

#### 4.3.2. Ważenie częstości termów

Inną metodą ekstrakcji cech jest wykorzystanie ważenia częstości termów (*TFIDF*) [25]. Jest to sposób obliczania wagi słów w oparciu o liczbę ich wystąpień w pojedynczym dokumencie oraz całym korpusie. Wartość *TFIDF* oblicza się z następującego wzoru:

$$(tf - idf)_{i,j} = tf_{i,j} \times idf_i$$

gdzie częstość termów wyrażana jest przez:

$$tf_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}}$$

a odwrotna częstość dokumentów:

$$idf_i = \log \frac{|D|}{|\{d : t_i \in d\}|}$$

gdzie:

$i, j$  - indeks słowa w słowniku oraz dokumentu

$|D|$  - liczba dokumentów w korpusie

$|\{d : t_i \in d\}|$  - liczba dokumentów zawierających przynajmniej jedno wystąpienie danego terminu

Przykładowo dla poniższego korpusu:

To jest pierwszy dokument  
Ten dokument jest drugim dokumentem  
I to jest trzeci z nich  
Czy to jest pierwszy dokument

reprezentacja kolejnych dokumentów w postaci wektorowej, z wykorzystaniem *TFIDF* i po lematyzacji, przedstawiono w tabeli 4.3.

L.p.	czy	dokument	drugi	jest	nich	pierwszy	to	trzeci
1	0	0,51	0	0,42	0	0,63	0,42	0
2	0	0,72	0,56	0,29	0,29	0	0,29	0
3	0	0	0	0,33	0,63	0	0,33	0,63
4	0,62	0,40	0	0,33	0	0,49	0,33	0

**Tabela 4.3.** Reprezentacja zdań przy pomocy metody *TFIDF*

Ekstrahowaną cechą jest wartość parametru *TFIDF* dla danego słowa w określonym dokumencie. Rozmiar wektora, dla każdego z nich równy więc będzie liczbie unikalnych słów w korpusie.

Przykładową implementację tego algorytmu, w języku *Python*, oferuje biblioteka *Sklearn*.

#### 4.3.3. Bag of words (BOW)

Kolejną metodą ekstrakcji cech jest wykorzystanie metody *bag of words* [26]. Każde unikalne słowo z korpusu zostaje dodane do worka (słownika) a następnie dla każdego dokumentu następuje weryfikacja, czy kolejne terminy z worka w nim występują. Tworzy się w ten sposób wektor zawierający cyfry 1 - dla pozytywnej weryfikacji - oraz 0 - dla negatywnej. Dla przykładu dokumenty z podrozdziału 4.3.2 reprezentowane są, w postaci wektorowej z wykorzystaniem *BOW* i lematyzacji, w sposób pokazany w tabeli 4.4.

#### 4. Klasyfikacja tekstów piosenek

---

L.p.	czy	dokument	drugi	jest	nich	pierwszy	to	trzeci
1	0	1	0	1	0	1	1	0
2	0	1	1	1	1	0	1	0
3	0	0	0	1	1	0	1	1
4	1	1	0	1	0	1	1	0

**Tabela 4.4.** Reprezentacja zdań przy pomocy metody *bag of words*

W tym wypadku ekstrahowaną cechą jest sama obecność terminu w dokumencie. Tak jak w przypadku *TFIDF* rozmiar pojedynczego wektora jest równy sumie wyrazów w słowniku.

### 4.4. Analiza wydźwięku

Mając zbiór danych - wyekstrahowany wektor cech razem z etykietami - można przejść do właściwej budowy modelu. Najpierw jednak należy podzielić dane na trzy podzbiory - treningowy, walidacyjny, zwykle w proporcji 8/1/1.

#### 4.4.1. Wybór klasyfikatora

Na tym etapie, należy wybrać klasyfikator jaki zostanie wykorzystany do klasyfikacji tekstów utworów. Najczęściej wykorzystywanymi, w tej problematyce, są:

- Maszyna wektorów nośnych (*SVM*)
- Naiwny klasyfikator Bayesa
- Konwulsyjne sieci neuronowe (*CNN*)

Maszyna wektorów nośnych (*SVM*) jest stosunkowo młodą metodą, zaproponowaną przez Vladimira Vapnika i Have Siegelmann w 1995 roku. Znajduje swoje zastosowania w problemach regresji oraz klasyfikacji [6]. Ta druga realizowana jest podobnie jak w przypadku sztucznych sieci neuronowych - przy pomocy zbioru uczącego. Istotą metody *SVM* jest konstrukcja optymalnej hiperpłaszczyzny, której zadaniem jest rozdzielenie danych, należących do przeciwnych klas z możliwie największym "marginesem zaufania"czyli odległością między danymi a hiperpłaszczyzną. Wykorzystuje się go w zadaniach wielowymiarowych, nawet takich w których liczba wymiarów jest większa niż próbek.

Naiwny klasyfikator Bayesa jest prostym klasyfikatorem probabilistycznym, cechującym się wysoką skutecznością co znajduje potwierdzenie w jego popularności przy rozwiązywaniu problemów klasyfikacji [27]. Naiwność wynika z założenia wzajemnej niezależności zmiennych niezależnych. Jego model wyprowadza się korzystając z twierdzenia



Bayesa. W tego typu klasyfikatorze wykorzystuje się różne rozkłady prawdopodobieństwa. Wyróżnić można:

- rozkład Gaussa
- rozkład Bernoulliego
- rozkład wielomianowy

Konwulsyjne sieci neuronowe (*CNN*) to jeden z typów głębokich sieci neuronowych, ze sprzężeniem zwrotnym [28][29]. Wykorzystywane głównie do klasyfikacji obrazów, osiągają jednak obiecujące wyniki także dla zastosowań analizy języka naturalnego.

### 4.4.2. Budowa modelu

Po wyborze klasyfikatora należy przejść do wyuczenia modelu, z wykorzystaniem wybranych wcześniej danych treningowych. Taki klasyfikator opisuje dzięki temu dane, którymi został wyuczony i ma możliwość dopasowywania nowych, nie widzianych jeszcze przez niego próbek, do zdefiniowanych klas.

Podczas budowy modelu, w zależności od wybranego klasyfikatora, do dyspozycji mamy różne parametry modyfikujące jego budowę. Dla przykładu w klasyfikatorze wykorzystującym *SVM* możliwy jest wybór jądra algorytmu [30]. Do dyspozycji są m.in.:

- liniowe
- wielomianowe
- gaussowskie
- sigmoidalne

Do zastosowań w prostych zadaniach klasyfikacji najczęściej wykorzystywane jest jądro liniowe. W przypadku innego wyboru, często konieczne jest zdefiniowanie innych parametrów takich jak kąt wielomianu.

### 4.5. Ewaluacja modelu

Gdy klasyfikator został już wyuczony, kolejnym krokiem jest jego ewaluacja czyli weryfikacja jakości klasyfikacji [31]. W praktyce stosuje się kilka metod, które zastosowane w kombinacji, pozwalają na obiektywne porównanie różnych modeli i wybranie najlepszego dla danego zastosowania. Do ewaluacji modelu wykorzystuje się dane, z wcześniej wydzielonego, zbioru testowego.

### 4.5.1. Dokładność

Podstawową, wykorzystywaną niemal zawsze, metodą jest obliczanie wartości dokładności (*ang. accuracy*), czyli stosunku liczby poprawnie zaklasyfikowanych próbek do ich całkowitej ilości. Przykładowo klasyfikator, który dla zbioru danych testowych zawierających 100 piosenek, poprawnie przydzielił klasę 70 utworom, charakteryzuje się dokładnością na poziomie 70%.

Parametr ten jednakże nie zawsze jest dobrą miarą jakości klasyfikacji. Przykładowo, dla niezbalansowanego zbioru testowego, zawierającego 85 utworów o wydźwięku *relaksującym* oraz po 5 o wydźwięku *smutnym*, *szczęśliwym* i *agresywnym*, klasyfikator przydzielający wszystkim piosenkom klasę *relaksująca* uzyska bardzo wysoką dokładność (85%). Wyniku tego nie można uznać jednak za wiarygodny. Aby uniknąć tego typu sytuacji wykonuje się balansowanie danych oraz korzysta z innych metod określania jakości klasyfikacji jak np. macierz błędów.

### 4.5.2. Macierz błędów

Inną metryką służącą do oceny jakości klasyfikacji jest macierz błędów inaczej nazywana tablicą pomyłek. Dla próbek z każdej klasy, przeprowadzana jest predykcja a jej wyniki - liczba próbek zakwalifikowanych do danej klasy - umieszczane są jako wiersz macierzy  $W$  ogólności wymiary macierzy zależne są więc od ilości klas do których przydzielane są próbki. Przykładową macierz błędów - dla modelu klasyfikującego piosenki do czterech grup - prezentuje tabela 4.5. Analizując jej zawartość można zauważyć, że najlepiej klasyfikowane są utwory o wydźwięku nazwanym jako *złość* natomiast najgorzej relaksujące oraz smutne.

	<b>Złość</b>	<b>Relaks</b>	<b>Szczęście</b>	<b>Smutek</b>
<b>Złość</b>	2372	74	32	47
<b>Relaks</b>	203	1661	225	431
<b>Szczęście</b>	129	289	1982	166
<b>Smutek</b>	284	527	180	1687

**Tabela 4.5.** Przykładowa macierz błędów, dla klasyfikatora, grupującego utwory na cztery kategorie

## 5. Analiza statystyczna melodii

W zadaniach analizy wydźwięku utworów instrumentalno-wokalnych, podstawą do analizy nie zawsze jest tekst. Współcześnie często wykorzystuje się w tym celu dane statystyczne charakteryzujące melodię piosenki [6]. Wynika to z braku konieczności zamiany danych na liczbowe - cechy statystyczne melodii są od razu w takiej formie - oraz dobrych wyników uzyskiwanych w ten sposób. Zdecydowaną wadą tego typu rozwiązań jest słaba dostępność danych. W przeciwieństwie do tekstów piosenek, sama melodia - zapisana cyfrowo lub nutowo - nie jest ogólnie dostępna, w szczególności dla utworów komercyjnych. Chroni ją szereg praw autorskich, które utrudniają dostęp do masowej ilości danych.

### 5.1. Cechy melodii

Melodię scharakteryzować można na wiele sposobów. Najczęściej określa się ją poprzez szereg elementów, z których każdy opisuje inny jej poziom, wymiar muzyki [32]. Poniżej przedstawiono zestawie najważniejszych cech melodii oraz terminów z nimi powiązanych.

#### 5.1.1. Rytm

Rytm jest elementem czasu w muzyce. Opisuje jak zmienia się utwór w przekroju czasu. Wyróżnić można kilka pojęć powiązanych z rytmem, z czego najważniejsze z nich to:

- tempo - jeden z najczęściej wykorzystywanych parametrów w analizie wydźwięku. Określa ilość *bitów* na minutę utworu (*bpm*). Czym wyższe tempo melodii tym większa jej dynamika. Najczęściej wysokim *bpm* charakteryzują się gatunki taneczne, R&B, Rock and Roll, natomiast niskim ballady, Blues oraz Jazz. Tempo ma znaczący wpływ na to jak odbierana jest melodia. Utwory o wysokim tempie najczęściej kojarzą się pozytywnie, w przeciwieństwie do piosenek o niskiej wartości tego parametru, które wywołują smutek, nostalgię ale i bywają relaksujące.
- czas trwania - określa czas trwania dźwięku (lub ciszy) w sekundach.
- metrum - jest to obowiązujący w utworze schemat, który określa układ akcentów w obrębie taktu, opisujących jego rytm.
- synkopa - pewne zaburzenia rytmu wykorzystywane w muzyce jako zabiegi stylistyczne.

## 5. Analiza statystyczna melodii

---

- ritardando - zwolnienia tempa w rytmie utworu.
- accelerando - przyspieszenia tempa w rytmie melodii.
- polirytm - występowanie więcej niż jednego rytmu jednocześnie.

### 5.1.2. Dynamika

Dynamika określa względną głośność lub cichość muzyki. Wysoką dynamiką charakteryzują się utwory Metalowe, Rockowe i Punkowe natomiast niską Bluesowe, Jazzowe i klasyczne. Poziom głośności wpływa w znaczący sposób na odbiór emocjonalny melodii - wysoki kojarzy się z agresją, natomiast niski z relaksem. Nie bez znaczenia jest także zmiana dynamiki w trakcie trwania utworu. Opisują to takie pojęcia jak *crescendo*, *più piano* i *più forte*.

### 5.1.3. Charakterystyka tonalna

Określa wysokość tonów w utworze. Wysokość dźwięku zależna jest od częstotliwości drgania instrumentów lub ich elementów (np. strun). Opis tonów w utworze - razem z ich długością i odległością między nimi - tworzy melodię.

### 5.1.4. Barwa

Barwa dźwięku inaczej nazywana tembrem jest to cecha dźwięku, odróżniająca brzmienie różnych instrumentów. Uzależniona jest od ilości, rodzaju i natężenia tonów składowych, ponieważ jest powiązane ze spektrum harmonicznym. Barwa dźwięku zawierającego silne tony jest wyrazista, w przeciwnym wypadku bezbarwna i głucha. O barwie decydują wzajemne stosunki amplitud poszczególnych tonów składowych, ich rozkład w widmie oraz jego struktura. Wyróżnia się barwy :

- alikwotowe - tony składowe mają równomierny rozkład w widmie
- formantowe - pewnie grupy tonów składowych są wzmocnione w widmie

## 5.2. Źródła charakterystyki statystycznej muzyki

Tak jak wcześniej wspomniano wszelkiego rodzaju zapisy cyfrowe oraz nutowe utworów nie są dostępne na dużą skalę. Ten problem rozwiązuje się wykorzystując serwisy streamingowe np. *Spotify*<sup>3</sup>, które udostępniają - zarejestrowanym użytkownikom, posiadającym konta premium - dostęp do gotowych, wydobytych cech statystycznych piosenek.

---

<sup>3</sup> <https://developer.spotify.com/>

Tego typu źródła danych dostarczają - poza podstawowymi cechami muzyki - także dodatkowe parametry, zaproponowane przez same serwisy. Zestaw cech udostępniany przez serwis streamingowy *Spotify* zawiera m.in. [33]:

- akustyczność - parametr określający prawdopodobieństwo, że określony utwór jest akustyczny. Bazuje na analizie rozpiętości tonalnej utworu. Duża ilość wysokich dźwięków zwiększa prawdopodobieństwo akustyczności. Parametr przyjmuje wartości od 0 do 1, gdzie 1 określa pewność, że utwór jest akustyczny.
- taneczność - określa jak dany utwór nadaje się do zastosowań tanecznych. Opiera się na złożeniu tempa, stabilności oraz siły rytmu. Przyjmuje wartości od 0 do 1.
- czas trwania - czas trwania utworu w milisekundach
- energia - parametry opisujący poziom energetyczności utworu. Pochodzi ze złożenia głośności, barwy i tempa. Wysoką energią charakteryzują się szybkie, dynamiczne utwory np. Rockowe. Przyjmuje wartości od 0 do 1.
- instrumentalność - określa prawdopodobieństwo czy dany utwór jest instrumentalny - nie posiada wokalu.
- klucz - wartość tonacji danego utworu z wykorzystaniem notacji klasy tonów (*Pitch class notation*)
- głośność - określa ogólną głośność utworu w dB. Wartość jest uśredniana dla całej ścieżki. Zwykle jest to wartość od -60 do 0 dB.
- tempo - parametr opisujący ogólne tempo utworu w bitach na minutę (*bpm*).
- sygnatura czasowa - szacowane metrum utworu, określa liczbę bitów w takcie.
- walencyjność - miara pozytywności piosenki. Przyjmuje od 1 do 0, gdzie wysoka wartość charakteryzuje utwory brzmiące pozytywnie (wesoło, euforycznie), natomiast niska negatywnie (agresywnie, smutno). Z punktu widzenia analizy wydźwięku jest to bardzo ciekawy parametr, ponieważ jest z nim bezpośrednio powiązany.

Powyżej przedstawione parametry są uśredniane dla całego utworu, jednakże dostępny jest także ich rozkład w czasie oraz szereg innych cech, mogących być bardzo dobrym źródłem danych dla wielu analiz. Sam serwis stale poszerza swoją ofertę oraz udoskonala już udostępnione informacje.

### **5.3. Klasyfikacja z wykorzystaniem cech statystycznych**

Proces klasyfikacji z wykorzystaniem cech statystycznych melodii nie różni się w sposób znaczący od tej wykonywanej przy pomocy tekstu [34]. Na pewien sposób można uznać ją nawet za jej uproszczoną wersję. Po pozyskaniu danych oraz ich zbalansowanie, nie ma potrzeby przeprowadzania serii operacji charakterystycznych dla danych tekstowych - tokenizacji, filtracji, ujednolicania słów. Często też jako wektor cech wykorzystuje się same cechy statystyczne danej melodii, nie wymagana jest więc, czasem skomplikowana i czasochłonna, ekstrakcja. Ważne jest aby z przedstawionych wcześniej cech melodii wybrać te, które wpływają w największym stopniu na wydźwięk utworu.

## 6. Projekt

Przedstawiony poniżej projekt magisterski koncentruje się na zbadaniu zależności między wydźwiękiem emocjonalnym tekstów utworów a - wymienionymi w rozdziale piątym - cechami statystycznymi melodii. Na projekt składa się zebranie danych (utworów, tekstów oraz opisu statystycznego), klasyfikacja wykorzystująca metody słownikowe, do jednej z czterech kategorii (złość, smutek, szczęście, relaks) a następnie zbadanie korelacji otrzymanych wyników z cechami melodii. Dodatkowo w ramach pracy, przeanalizowano jak często poszczególne cechy występują w zależności od wydźwięku emocjonalnego utworów.

Implementacja wykonywana była w języku *Python 3.7* z wykorzystaniem bibliotek i narzędzi, które szczegółowo opisano w podrozdziale 6.6.

### 6.1. Źródła danych

W trakcie wykonywania projektu magisterskiego korzystano z szeregu źródeł danych. Poniżej przedstawiono ich zestawienie.

#### 6.1.1. Million Song Dataset

*Million Song Dataset* jest darmowym zbiorem zawierającym informacje o około milionie utworów muzycznych. *MSD* dostępny jest w dwóch formach - plików tekstowych oraz bazy danych *sqlite* [35]. Każdy rekord (utwór) opisywany jest przez szereg pól, z których najważniejsze to:

- nazwa twórcy
- identyfikator twórcy w serwisie *musicbrainz.org*
- identyfikator twórcy w serwisie *Echo Nest*
- nazwa utworu
- rok produkcji
- nazwa albumu
- identyfikator utworu w serwisie *Echo Nest*

Taki zbiór danych - z punktu widzenia praktycznego zastosowania w analizie wydźwięku - jest zbyt duży. Operacje na nim wykonywane trwały by zbyt długo. W związku z tym skorzystano z losowego podzbioru zawierającego 1% rekordów czyli około dziesięciu tysięcy utworów.

### 6.1.2. Lyric Wikia

Warto zauważyć, że przedstawione wcześniej źródło danych nie zawiera tekstów utworów - powodem są prawa autorskie. W związku z tym skorzystano z biblioteki *PyLyrics*, która pozwala na ich pobranie dla wskazanego utworu - identyfikowanego przez tytuł i nazwę autora [36]. Źródłem informacji dla powyższego pakietu jest serwis Lyric Wikia<sup>4</sup>, w którym każdy użytkownik może umieścić tekst dla dowolnej piosenki.

Na tym etapie przeprowadzono też wstępną filtrację utworów. Usunięto te, dla których biblioteka *PyLyrics* nie znalazła tekstu. Warto zauważyć, że już na tym etapie pojawiają się pewnie zagrożenia. Z uwagi na społecznościowy charakter źródła tekstów istnieje szansa, że będą one niepoprawne. Dodatkowo nie ma pewności, iż identyfikator budowany na podstawie tytułu oraz nazwy artysty jest unikalny.

### 6.1.3. Spotify API

Źródłem cech statystycznych było *API*<sup>5</sup> serwisu *Spotify*. Ich skład przedstawiono już w podrozdziale 5.2. *Endpoint*<sup>6</sup> udostępniający poszukiwane dane wymaga identyfikatora utworu - jako parametru - dlatego wcześniej należało go pozyskać z innego *endpoint'a*. Przykładową odpowiedź przedstawia rysunek 6.1.

---

<sup>4</sup> <https://lyrics.fandom.com/wiki/LyricWiki>

<sup>5</sup> API (ang. Application Programming Interface) - protokół, umożliwiający dwóm aplikacjom komunikację między sobą

<sup>6</sup> Endpoint - adres pod którym *API* udostępnia określone dane



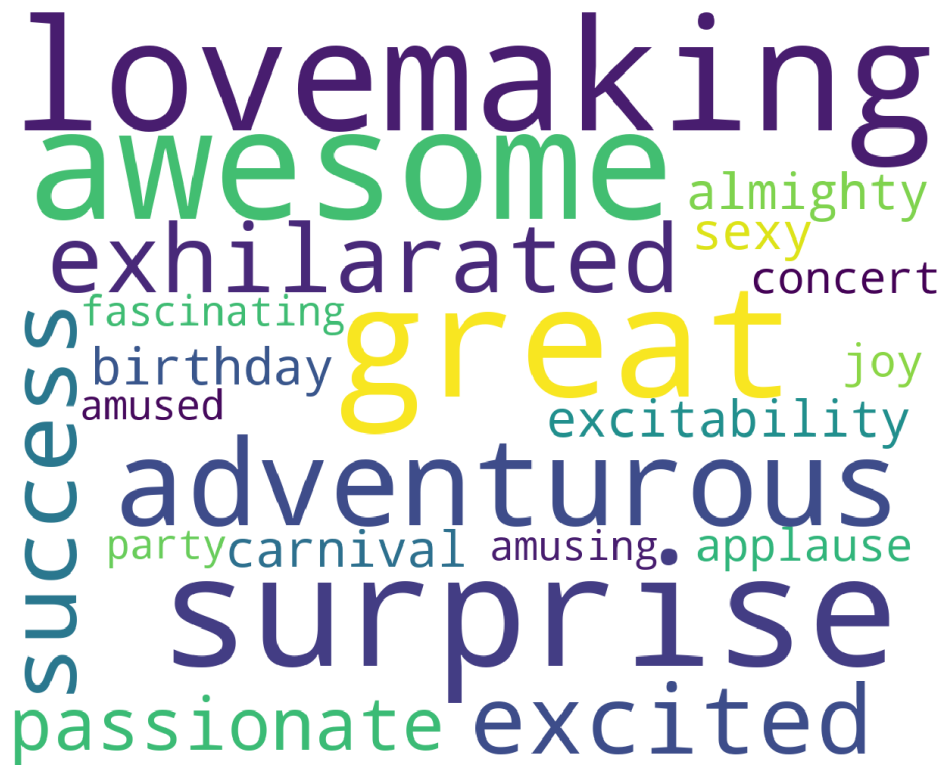
```
{
  "duration_ms" : 255349,
  "key" : 5,
  "mode" : 0,
  "time_signature" : 4,
  "acousticness" : 0.514,
  "danceability" : 0.735,
  "energy" : 0.578,
  "instrumentalness" : 0.0902,
  "liveness" : 0.159,
  "loudness" : -11.840,
  "speechiness" : 0.0461,
  "valence" : 0.624,
  "tempo" : 98.002,
  "id" : "06AKEBrKUckWOKREUWRnvT",
  "uri" : "spotify:track:06AKEBrKUckWOKREUWRnvT",
  "track_href" :
  "api.spotify.com/v1/tracks/06AKEBrKUckWOKREUWRnvT",
  "analysis_url" :
  "api.spotify.com/v1/audio-analysis/06AKEBrKUckWOKREUWRnvT",
  "type" : "audio_features"
}
```

**Rysunek 6.1.** Przykładowa odpowiedź API serwisu *Spotify* z cechami utworu

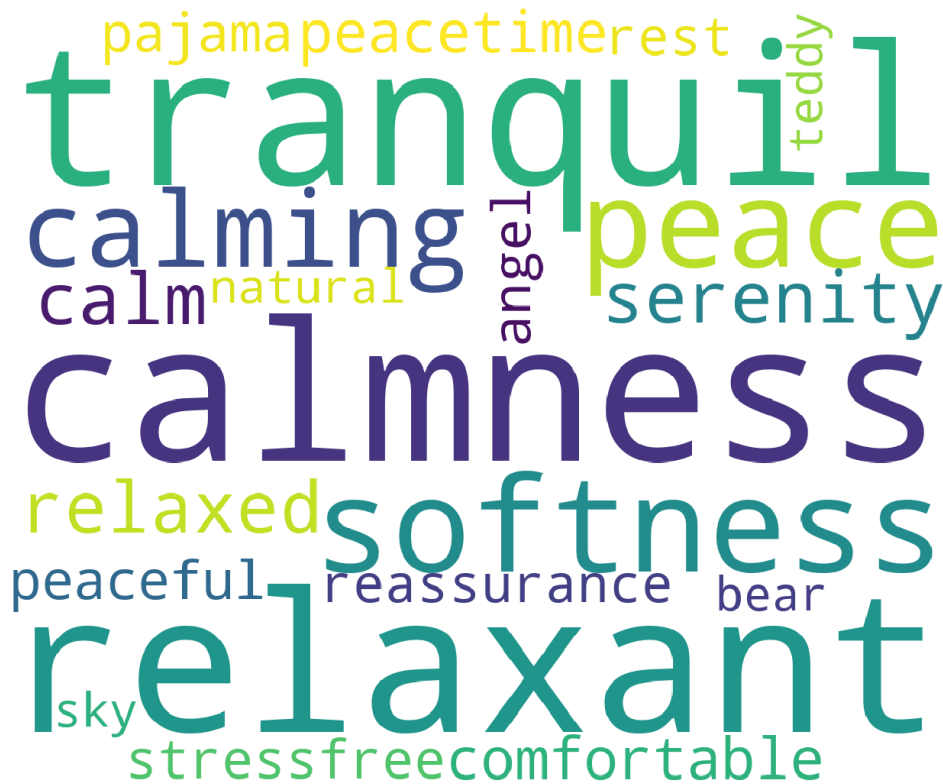
#### 6.1.4. Leksykon *NRC*

W celu klasyfikacji konieczne było skorzystanie z leksykonu emocjonalnego - słownika zawierającego, dla każdego ze słów, wydźwięk jaki ze sobą niesie. Dostępnych jest wiele leksykonów - w szczególności dla języka angielskiego - różniących się typem zwracanych wartości. Większość z nich oferuje rozróżnienie wydźwięku jedynie na pozytywny i negatywny, część w sposób liczbowy np. od -5 do 5. Wybrany leksykon *NRC* udostępnia także m.in. poziom *walencyjności* oraz *pobudzenia* [37], dla każdego z terminów, co pozwala na wykorzystaniu go w zagadnieniu klasyfikacji do klas z modelu *Russella*.

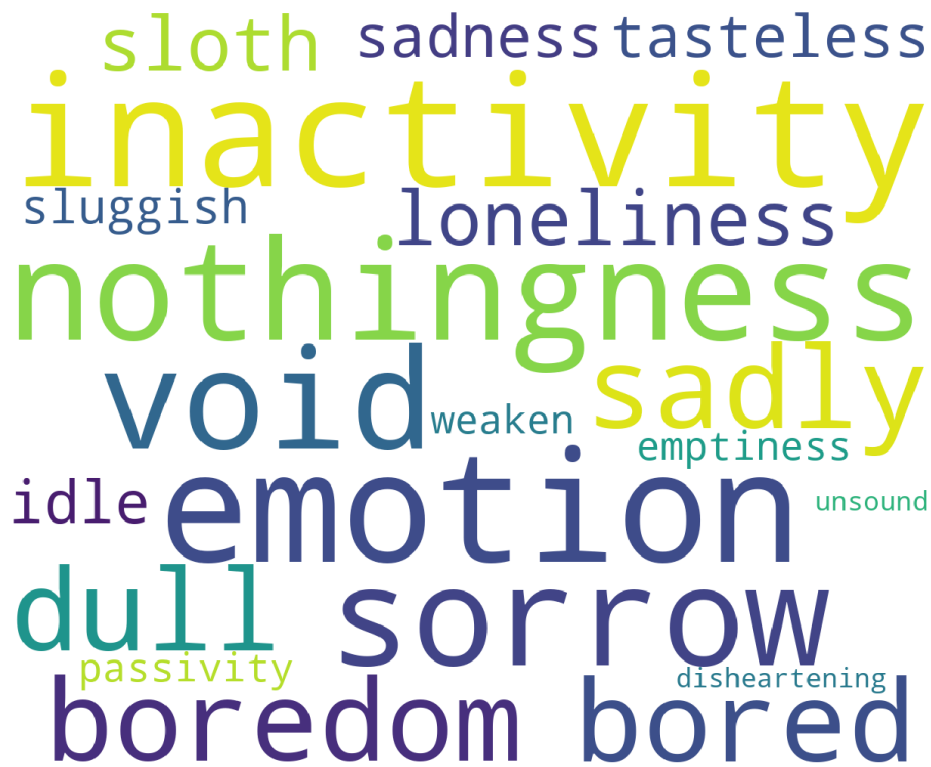
Na rysunkach 6.2, 6.3, 6.4, 6.5 przedstawiono tzw. *word cloud*'y - wizualizacje wagi słów, w zależności od określonego parametru - prezentujące wyrazy charakteryzujące się kolejno *szczęśliwym*, *relaksującym*, *smutnym* oraz *gniewnym* wydźwiękiem, według leksykony *NRC*



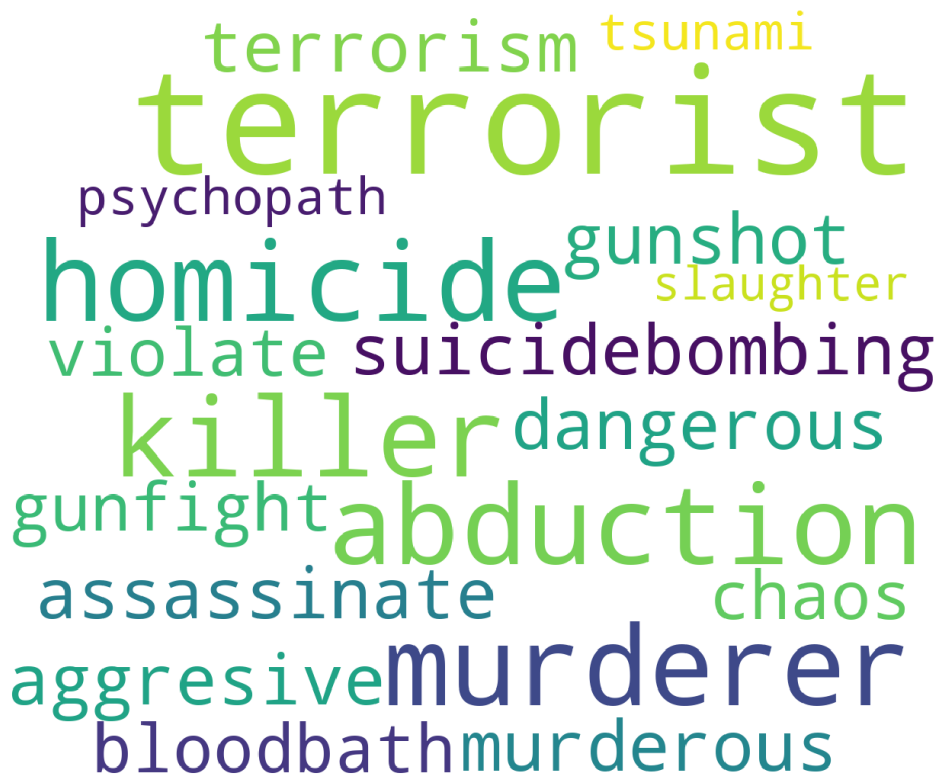
**Rysunek 6.2.** Word cloud prezentujący słowa charakteryzujące się *szczęśliwym* wydźwiękiem, według leksykonu NRC



**Rysunek 6.3.** Word cloud prezentujący słowa charakteryzujące się *relaksującym* wydźwiękiem, według leksykonu NRC



**Rysunek 6.4.** *Word cloud* prezentujący słowa charakteryzujące się *smutnym* wydźwiękiem, według leksykonu NRC



**Rysunek 6.5.** *Word cloud* prezentujący słowa charakteryzujące się *gniewnym* wydźwiękiem, według leksykonu NRC

### 6.1.5. Tagi *last.fm*

Aby zweryfikować uzyskane wyniki skorzystano z tagów serwisu *last.fm*, umożliwiającemu użytkownikom podzielenie się emocjami towarzyszącym słuchaniu danego utworu. Otrzymanie etykiet wymaga obróbki uzyskanych tagów - filtracja tych, które określają emocje, wyznaczenie dominującej z nich - szczęśliwie, z uwagi na swoją popularność, w analizie wydźwięków, istnieją gotowe zbiory, wyznaczonych już etykiet.

## 6.2. Analiza przy użyciu metody słownikowej

Na tym etapie warto wspomnieć, że wybranie metody słownikowej do klasyfikacji utworów nie jest przypadkowe. Nie wykorzystuje ona, w przeciwieństwie do metod statystycznych z wykorzystaniem nadzorowanego uczenia maszynowego, etykiet, do których uzyskania korzysta się z opinii użytkowników, bazujących na całokształcie utworu - tekście oraz melodii. Takie podejście gwarantuje pełną niezależność klasyfikacji od ścieżki melodycznej.

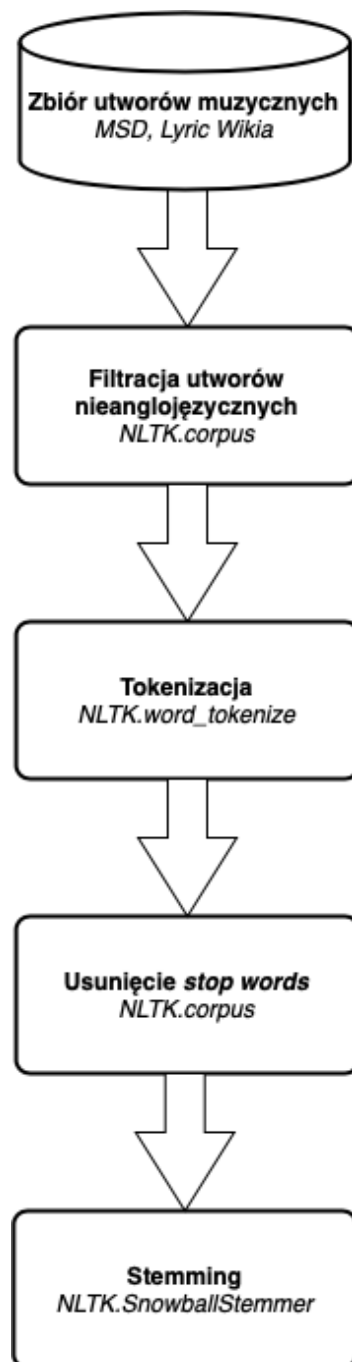
Proces analizy bazującej na metodzie słownikowej podzielić można na szereg etapów. Przedstawiono je na rysunku 6.6.



**Rysunek 6.6.** Proces analizy przy pomocy metody słownikowej wykorzystującej leksykon

### 6.2.1. Obróbka danych

Tak jak wspomniano w podrozdziale 4.2 dane wymagają wstępnej obróbki aby zapewnić wysoką jakość klasyfikacji. Rysunek 6.7 prezentuje wykonane czynności.



Rysunek 6.7. Kolejne etapy obróbki danych

Z uwagi na fakt, że klasyfikacja wykorzystywać będzie leksykon anglojęzyczny, konieczne jest odfiltrowanie utworów, których cały tekst lub jego fragment jest w innym języku. W tym celu wykorzystano pakiet *corpus* biblioteki *NLTK*, który zawiera spis wszystkich angielskich słów.

Po odfiltrowaniu części utworów wykonano tokenizację tekstów na pojedyncze słowa. Użyto w tym celu funkcję *word\_tokenize*, która przyjmuje dokument i zwraca go w formie listy tokenów.

Następnym krokiem było usunięcie tzw. *stop words*. W tym celu, dla każdego ze słów z korpusu sprawdzano przynależność do listy *stop\_words* z pakietu *NLTK*. W przypadku pozytywnego wyniku termin usuwano.

Kolejnym etapem obróbki danych był *stemming*, czyli wydobycie rdzenia ze słowa. Operację wykonano przy pomocy pakietu *SnowballStemmer*.

### 6.2.2. Ekstrakcja cech

Ekstrakcję cech wykonano przy pomocy algorytmu *TFIDF* - ważenia częstości termów. Użyto pakietu *TfidfVectorizer* z biblioteki *Sklearn*. Warto zauważyć, że pakiet ten pozwala na automatyczne przeprowadzenie wcześniejszej obróbki - tokenizacji, filtracji *stop words* - ale dla większej kontroli zachodzących procesów, nie skorzystano z tej opcji. Wycinek uzyskanych wektorów cech przedstawia rysunek 6.8.

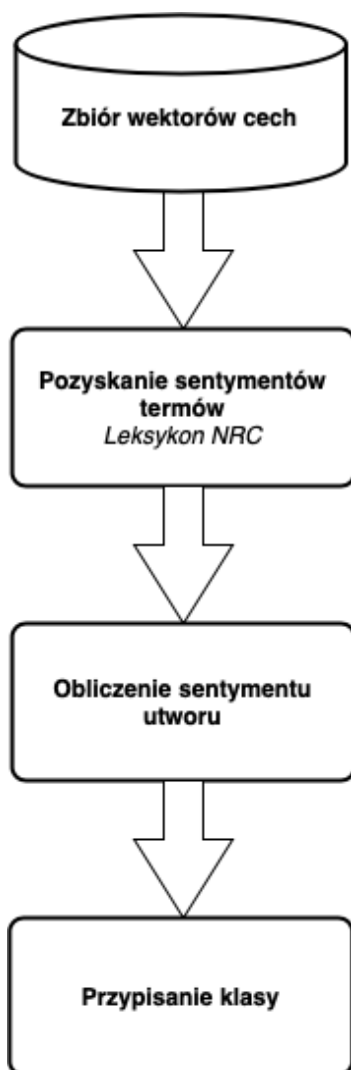
```
(0, 2095)      0.09370966352607901
(0, 8616)     0.09258204204640863
(0, 80)       0.07907019709227531
(0, 8600)     0.1092110823397121
(0, 2919)     0.04262245896119218
(0, 5789)     0.04465009073011278
(0, 8161)     0.06923043691018785
(0, 4638)     0.024318326799931955
(0, 2835)     0.04262245896119218
(0, 2445)     0.024884158988990903
(0, 3572)     0.029617071319389718
(0, 9352)     0.024335645525374105
(0, 5724)     0.05238791758614052
(0, 9646)     0.10446466120807567
(0, 3593)     0.16725101160829217
(0, 4943)     0.14962260955671325
(0, 6683)     0.09982652253111615
(0, 1139)     0.06743112112366692
(0, 9392)     0.07655285421778396
(0, 4184)     0.06034475971088869
(0, 5324)     0.0771599690112315
(0, 9596)     0.10852774157853691
(0, 3358)     0.04950504008580763
(0, 1819)     0.0342201429254578
(0, 8689)     0.027102887146694468
```

Rysunek 6.8. Wycinek wyniku algorytmu ważenia częstości termów dla danych projektowych

W lewej kolumnie widać numer dokumentu (piosenki) - iteracja rozpoczyna się od 0 - oraz numer słowa w słowniku, natomiast w prawej wartość *TFIDF*.

### 6.2.3. Klasyfikacja

Proces klasyfikacji składał się z kilku etapów, które - dla większej czytelności - zostały przedstawione na rysunku 6.9.



**Rysunek 6.9.** Proces klasyfikacji utworów muzycznych z wykorzystaniem leksykonu

Dla każdego słowa w danym utworze pobierano wartość wydziwki ze słownika. W tym celu skorzystano z leksykonu *NRC*. Słownik ten zwraca wartość *walencyjności* oraz *pobudzenia* dla każdego z terminów posiadających wydziwkę emocjonalny [37]. Uzyskane wartości liczbowe zostają przemnożone przez wagę słowa - pochodzącą z wektora cech - i zsumowane dla całego dokumentu. W ten sposób otrzymano złożenie wydziwki dla kolejnych utworów w postaci wektora dwuelementowego, gdzie pierwsza liczba określa



walencyjność, a druga *pobudzenie*. Poniżej przedstawiano powyższą metodę obliczeń dla kolejno *walencyjności* oraz *pobudzenia* w formie wzoru.

$$v_j = \sum_i (tf - idf)_{i,j} * v_{NRC_i}$$

$$a_j = \sum_i (tf - idf)_{i,j} * a_{NRC_i}$$

gdzie:

$i, j$  - indeks słowa oraz utworu

$(tf - idf)_{i,j}$  - wartość *TFIDF* dla słowa  $i$  w utworze  $j$

$v_{NRC_i}, a_{NRC_i}$  - wartość *walencyjności* oraz *pobudzenia* dla słowa  $i$  w leksykonie *NRC*

Zgodnie z wykresem modelu *Russella* przedstawionym na rysunku 3.1 wartość wydźwięku utworu jest uzależniona od znaku *walencyjności* oraz *pobudzenia*. Przykładowo dla utworu, którego  $v_j = 0,35$  oraz  $a_j = -0,11$  wydźwięk emocjonalny będzie *agresywny*. Odległość punktu w przestrzeni *walencyjno-pobudzeniowej* od początku układu współrzędnych określać będzie siłę wydźwięku.

#### 6.2.4. Wyniki

Wyniki klasyfikacji z wykorzystaniem metody słownikowej - liczba utworów przydzielonych do danych klas - przedstawiono w tabeli 6.1.

Relaks	Szczęście	Smutek	Złość
2415	2501	3172	1810

**Tabela 6.1.** Liczba utworów zakwalifikowanych do danej kategorii

W badanym podziorze najwięcej było utworów zaklasyfikowanych jako *smutne* natomiast najmniej *agresywnych*. Po sumarycznej liczbie zaklasyfikowanych utworów widać, że część z nich została usunięta w procesie obróbki - brak lub nieanglojęzyczny tekst.

#### 6.2.5. Ewaluacja rozwiązania

Do ewaluacji modelu skorzystano z wzorcowych etykiety uzyskanych z serwisu *last.fm*. Dokładność obliczono przy pomocy funkcji *accuracy\_score* z biblioteki *sklearn*, który na wejściu przyjmuje listę predykowanych oraz oczekiwanych klas, a na wyjściu podaje skuteczność predykcji w procentach. Dokładność dla tego rozwiązania wynosi **57.32%**. Macierz błędów przedstawiono w tabeli 6.2.

	<b>Relaks</b>	<b>Szczęście</b>	<b>Smutek</b>	<b>Złość</b>
<b>Relaks</b>	1468	220	342	385
<b>Szczęście</b>	377	1528	252	344
<b>Smutek</b>	332	479	1774	587
<b>Złość</b>	341	240	316	913

**Tabela 6.2.** Macierz błędów dla klasyfikatora słownikowego wykorzystującego leksykon *NRC*

Powyższe wyniki wskazują, że klasyfikacja cechowała się względnie wysoką dokładnością. Żaden z wydźwięków nie wyróżniał się jeśli chodzi o poprawność predykcji. Pamiętać należy, że wykorzystane etykiety odnoszą się do całościowego odbioru utworu a nie jedynie do tekstu, dlatego "prawdziwa" dokładność może nieco odbiegać od tej otrzymanej eksperymentalnie.

### 6.3. Analiza związku cech statystycznych melodii z wynikami klasyfikacji

Zaprezentowane w podrozdziale 6.2 wyniki wykorzystano do przeprowadzenia analizy związku wydźwięku tekstów utworów z cechami statystycznymi melodii. Wykonano szereg analiz przedstawionych poniżej.

#### 6.3.1. Badanie korelacji

Jedną z metod badania związku pomiędzy zmiennymi jest wyznaczanie współczynnika korelacji *Pearsona*. W tym celu skorzystano z funkcji *pearsonr* z biblioteki *SciPy*. Metoda przyjmuje dwie listy zmiennych, między którymi określana będzie korelacja. W tym przypadku jest to tablica wartości składowych wydźwięku emocjonalnego utworu - kolejno walencyjność oraz pobudzenie - oraz lista wartości wybranej cechy statystycznej melodii. Tabela 6.3 oraz 6.4 przedstawia wyniki tej operacji, dla kolejnych cech. Przyjęty poziom ufności wynosił 95%.

Cecha	Współczynnik korelacji	Dolna granica ufności	Górna granica ufności
energia	-0.29328740	-0.35434051	-0.24472147
tempo	-0.04915165	-0.08504889	-0.00275848
walencyjność	0.21121758	0.16531446	0.25770111
taneczność	0.04502676	-0.01854266	0.09104836
głośność	-0.26471991	-0.30962743	-0.21911894
akustyczność	0.01904423	0.01452150	0.02348754

**Tabela 6.3.** Współczynnik korelacji *Pearsona* między *walencyjnością* a kolejnymi cechami statystycznymi melodii

Cecha	Współczynnik korelacji	Dolna granica ufności	Górna granica ufności
energia	0.26643008	0.22319320	0.31257018
tempo	0.03687758	-0.02010128	0.07283128
walencyjność	0.24940954	0.2029025	0.29570322
taneczność	0.01448170	-0.03171127	0.06129558
głośność	0.29691072	0.25178732	0.34121505
akustyczność	-0.02899489	-0.04319852	-0.01467632

**Tabela 6.4.** Współczynnik korelacji *Pearsona* między *pobudzeniem* a kolejnymi cechami statystycznymi melodii

## 6. Projekt

Powyższe wyniki wskazują - w ramach przedziału ufności - występowanie korelacji między wydźwiękiem emocjonalnym utworów a *energią*, *głośnością* oraz *walencyjnością* - jako cechą statystycznej melodii. W przypadku dwóch pierwszych cech wykazano dodatnią korelację z *pobudzeniem* oraz ujemną z *walencyjnością*. Oznacza to, że np. utwory o wysokiej wartości *głośności* powinny być często klasyfikowane jako *agresywne*. W przypadku *walencyjności* korelacja w obydwu przypadkach była dodatnia, a więc wzrost tej cechy idzie w parze ze wzrostem *walencyjności* oraz *pobudzenia* tekstu utworu. Korelacja między resztą cech statystycznych a wydźwiękiem utworów jest znikoma lub wcale jej nie ma - zmienne są liniowo niezależne.

Pamiętać należy, że współczynnik *Pearsona* wyznacza jedynie liniową korelację. Metodą pozwalającą na zbadanie także nieliniowych zależności jest korelacja *Spearmana*. W tabeli 6.5 oraz 6.6 przedstawiono wartości współczynnika korelacji *Spearmana* między wydźwiękiem utworów a kolejnymi cechami statystycznymi melodii.

Cecha	Współczynnik korelacji	Dolna granica ufności	Górna granica ufności
energia	-0.27871424	-0.32337107	-0.23330876
tempo	-0.05116484	-0.09709713	-0.00484136
walencyjność	0.25118678	0.20861710	0.30812301
taneczność	0.04507438	-0.01861710	0.09125127
głośność	-0.17545715	-0.22017391	-0.13000459
akustyczność	0.03098369	-0.08430122	0.07246629

**Tabela 6.5.** Współczynnik korelacji *Spearmana* między *walencyjnością* a kolejnymi cechami statystycznymi melodii

Cecha	Współczynnik korelacji	Dolna granica ufności	Górna granica ufności
energia	0.21331528	0.16847326	0.25727620
tempo	0.05687758	-0.00445723	0.08840574
walencyjność	0.24940954	0.20506431	0.29784473
taneczność	0.01985014	-0.02668812	0.066302558
głośność	0.21064298	0.16575332	0.25466145
akustyczność	-0.02827685	-0.07500746	0.02394036

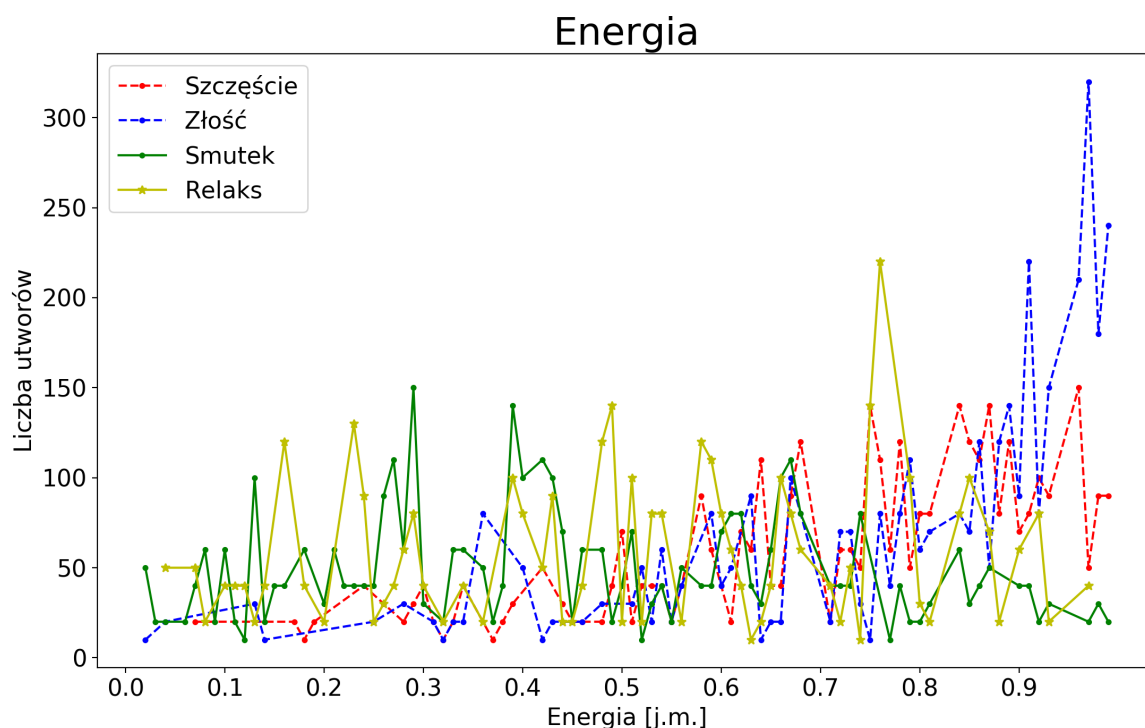
**Tabela 6.6.** Współczynnik korelacji *Spearmana* między *pobudzeniem* a kolejnymi cechami statystycznymi melodii

Z powyższych rezultatów wynika, że dla nieskorelowanych liniowo cech statystycznych - z wydźwiękiem utworów - nie występuje także żadna inna monotoniczna zależność.

### 6.3.2. Skorelowane cechy statystyczne a wydźwięk utworów

Istotnym elementem badań jest sprawdzenie - dla cech statystycznych uznanych za skorelowane - jak wydźwięk emocjonalny tekstu wpływa na cechy statystyczne melodii. Przykładowo czy utwory o *agresywnym* tekście częściej cechują się wysoką głośnością niż inne.

Na rysunku 6.10 przedstawiono zależność liczby utworów od średniej *energii* w nich występujących.

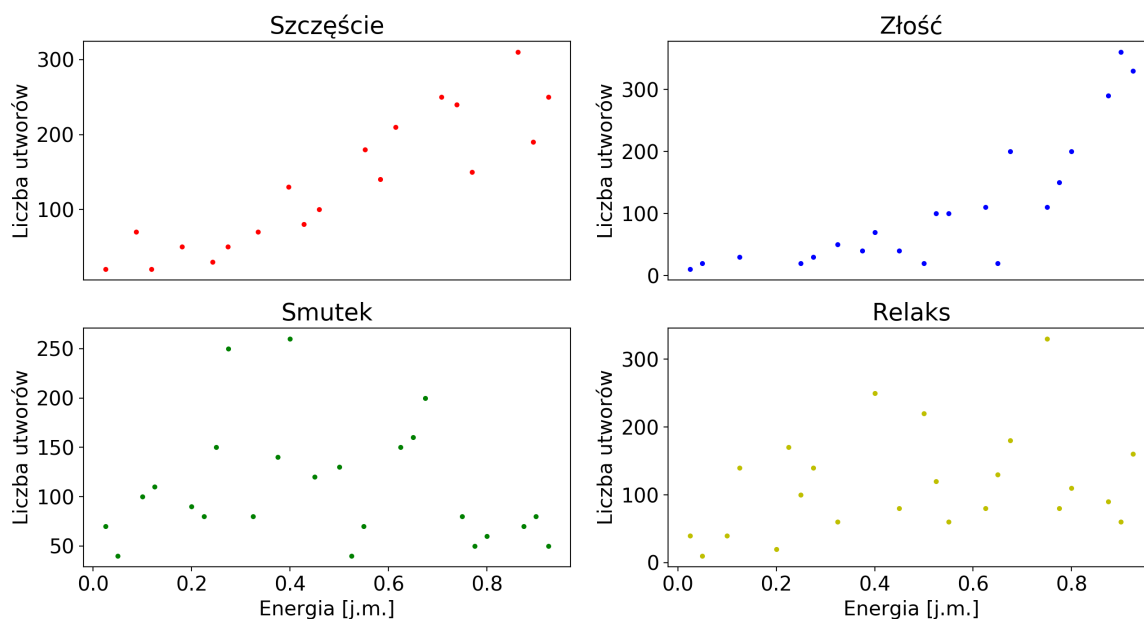


**Rysunek 6.10.** Zależność liczby utworów - o określonym wydźwięku - od średniej wartości energii

Na powyższym wykresie można zauważyć, że utwory cechujące się *gniewnym* wydźwiękiem występują najczęściej z dużą wartością średniej energii. W zakresie 0 – 0,3 j.m. występuje znikoma ilość tego typu piosenek. Zdecydowanie większość z nich znaleźć można między 0,8 – 1 j.m. Podobnie sytuacja wygląda dla utworów zdefiniowanych jako *szczęśliwe*, z tym że zakres występowania większości z nich to 0,6 – 1 j.m. W przypadku rozkładu piosenek *smutnych* oraz *relaksujących* tendencja nie jest jednoznaczna. Utwory rozkładają się dość równomiernie na całym przedziale wartości *energii*.

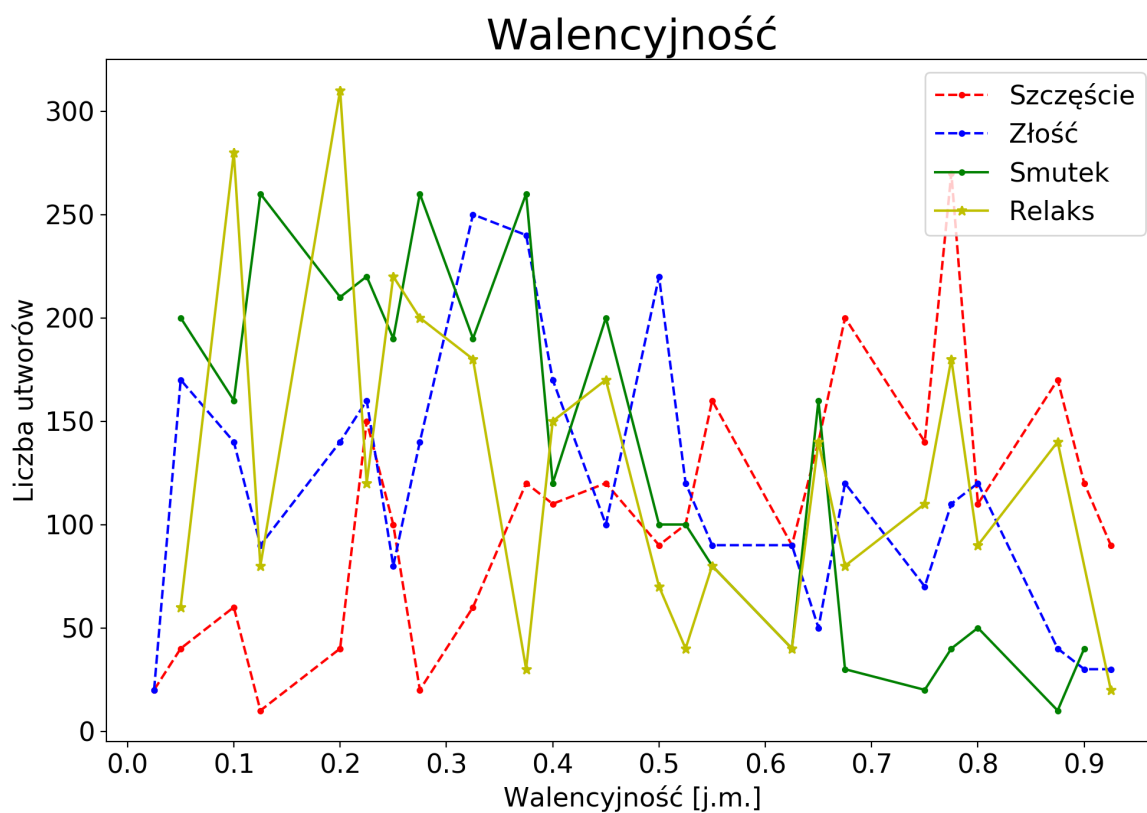
Na rysunku 6.11 przedstawiono te same dane, z podziałem na wydźwięk w celu lepszej prezentacji zależności.

## 6. Projekt



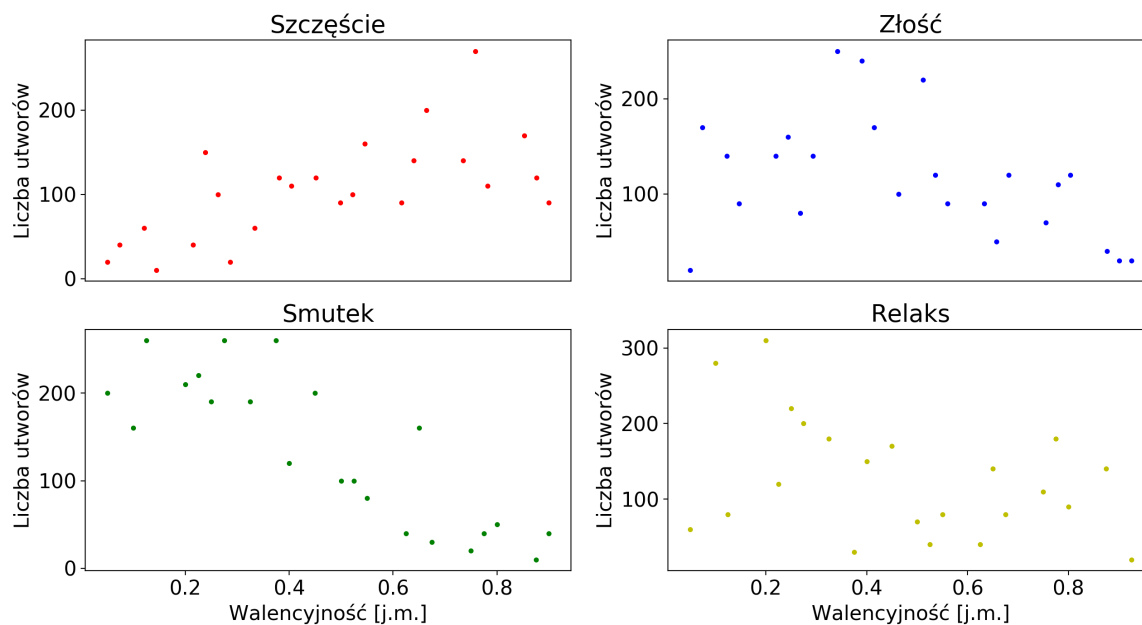
**Rysunek 6.11.** Zależność liczby utworów od średniej wartości energii dla piosenek kolejno *szczęśliwych, agresywnych, smutnych i relaksujących*

Zależność ilości utworów od średniej wartości *walencyjności* przedstawiono na rysunku 6.12.



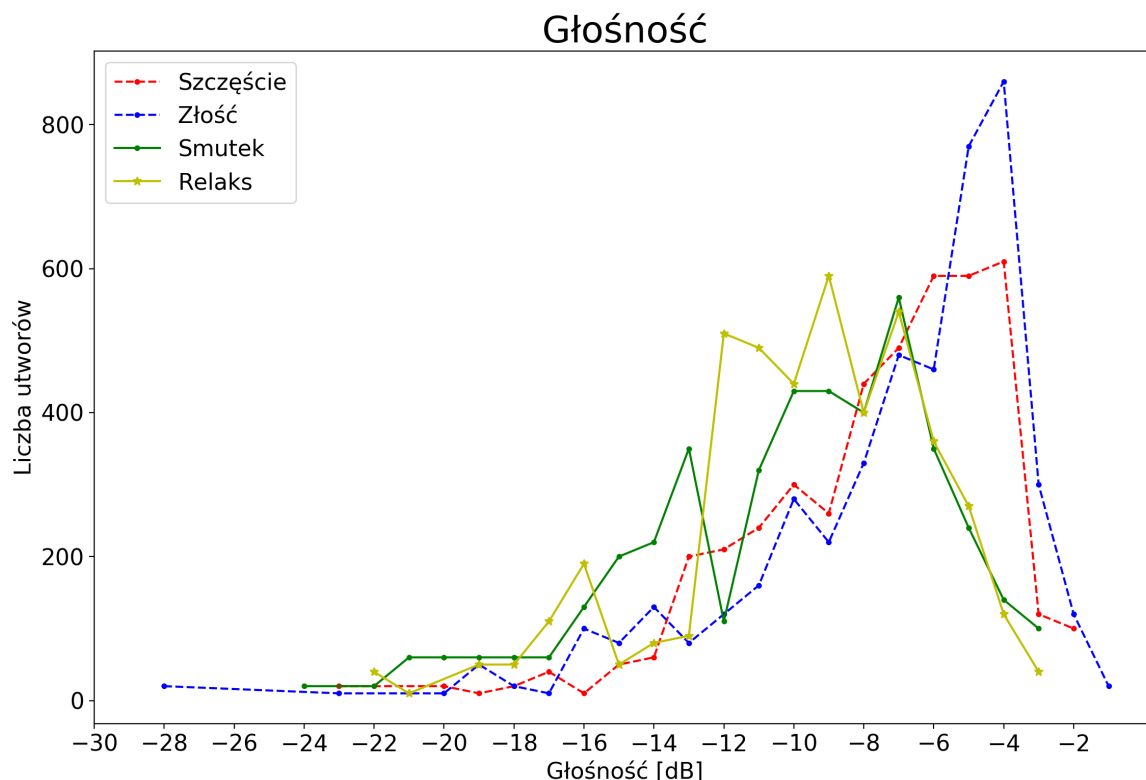
**Rysunek 6.12.** Zależność liczby utworów - o określonym wydźwięku - od średniej wartości walencyjności

Większość utworów oznaczonych jako *relaksujące* charakteryzują się niską wartością walencyjności (między 0 a 0,4 j.m.). Podobnie sytuacja wygląda dla *smutnych* (0 – 0,5 j.m.) oraz *agresywnych* (0 – 0,55 j.m.). W przypadku piosenek o *szczęśliwym* można dostrzec pewną tendencję wzrostu liczby utworów wraz ze wzrostem *walencyjności*. Poniżej na rysunku 6.13 przedstawiono zależność liczby utworów od od średniej wartości *walencyjności* z wyszczególnieniem poszczególnych wydźwięków.



**Rysunek 6.13.** Zależność liczby utworów od średniej wartości walencyjności dla piosenek kolejno *szczęśliwych*, *agresywnych*, *smutnych* i *relaksujących*

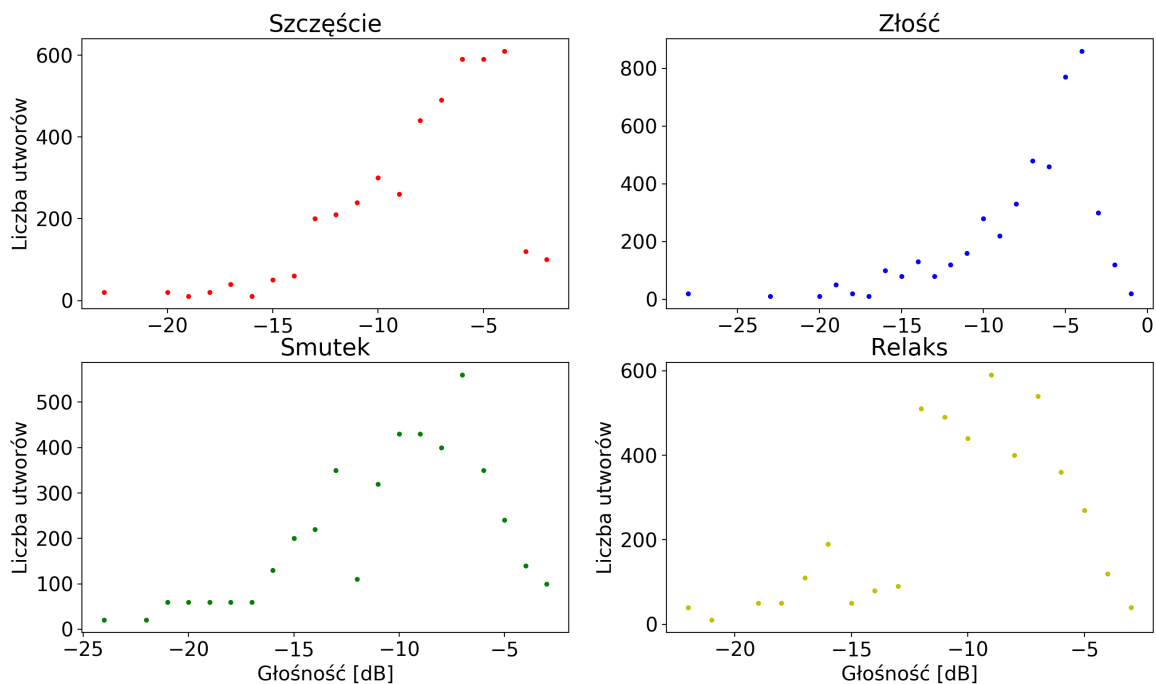
Zależność liczby utworów od średniej wartości *głośności* przedstawiono na rysunku 6.14.



**Rysunek 6.14.** Zależność liczby utworów - o określonym wydźwięku - od średniej wartości głośności

Zdecydowana większość zarówno utworów *szczęśliwych* jak i *złych* cechuje się wysoką średnią głośnością (-10 – -3 dB). *Relaksujące* piosenki najczęściej -14 – -5 dB natomiast *smutne* między -11 a -6 dB. Oznacza to, że utworom o tekście charakteryzującym się wysoką wartością *pobudzenia* (*szczęśliwy, agresywny*) towarzyszy zazwyczaj głośniejsza melodia. Dla niskiej wartości *pobudzenia* (*smutek, relaks*) melodia jest wyraźnie cichsza. Poniżej na rysunku 6.15 przedstawiono zależność liczby utworów od od średniej wartości *głośności* z wyszczególnieniem poszczególnych wydźwięków.



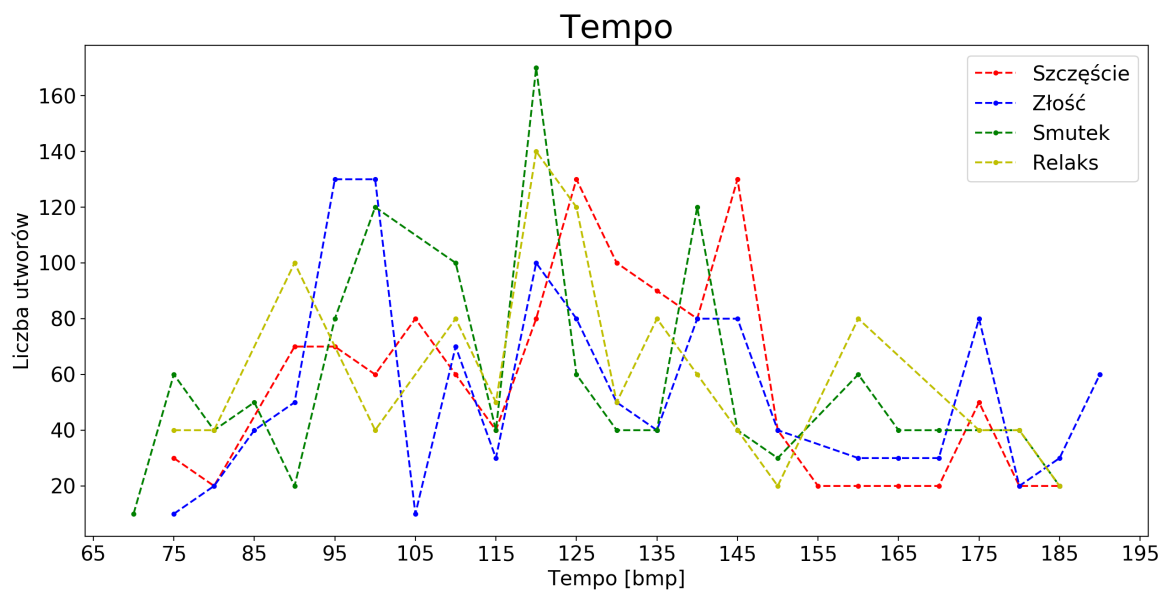


**Rysunek 6.15.** Zależność liczby utworów od średniej wartości głośności dla piosenek kolejno szczęśliwych, agresywnych, smutnych i relaksujących

### 6.3.3. Weryfikacja nieskorelowanej cechy

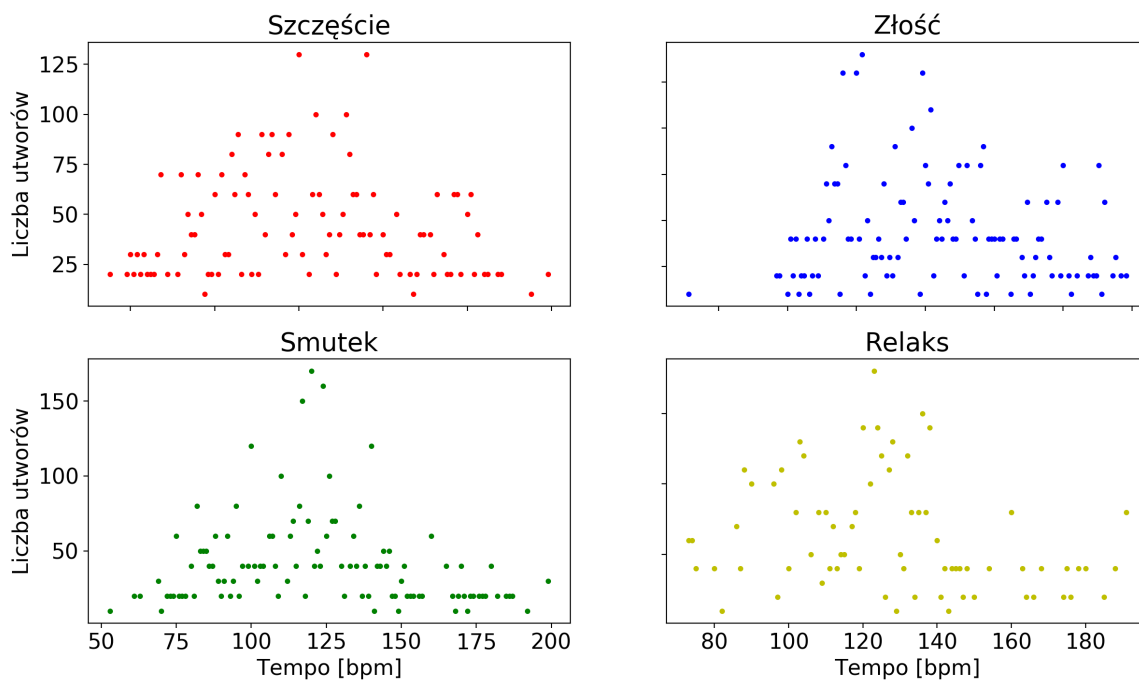
Dużym zaskoczeniem jest brak korelacji *tempa* utworu z wydźwiękiem emocjonalnym tekstu. Mogłoby się wydawać, że utwory, których słowa cechujące się większym *pobudzeniem* (*agresywne* oraz *szczęśliwe*) powinny charakteryzować się, także wyższym tempem melodii. W celu sprawdzenia czy na pewno korelacja nie występuje zbadano zależność liczby utworów od średniej wartości *tempa*, którą przedstawiono na rysunku 6.16.

## 6. Projekt



**Rysunek 6.16.** Zależność liczby utworów - o określonym wydźwięku - od średniej wartości tempa

Powyższy wykres potwierdza brak powiązania między tempem utworów a liczbą ich wystąpień dla poszczególnych wydźwięków. Ciężko doszukać się, dla którejś z emocji, tendencji wraz ze wzrostem lub spadkiem tempa. Poniżej na rysunku 6.17 przedstawiono zależność liczby utworów od od średniej wartości *tempa* z wyszczególnieniem poszczególnych wydźwięków.



**Rysunek 6.17.** Zależność liczby utworów od średniej wartości tempa dla piosenek kolejno *szczęśliwych, agresywnych, smutnych i relaksujących*

## 6.4. Podsumowanie wyników

Skonstruowany klasyfikator cechuje się - w porównaniu do innych tego typu rozwiązań - dość dobrą jakością klasyfikacji, dokładność - 57.32%. Analizując precyzyjność klasyfikatorów tekstów piosenek, można zauważyć, że najlepszymi rezultatami mogą pochwalić się te bazujące na nadzorowanym uczeniu maszynowym (wyniki na poziomie 70% poprawności predykcji), jednakże ze względu na pochodzenie etykiet (tagi *last.fm*) wykorzystanie ich jako docelowe rozwiązanie nie było wskazane. Macierz błędów potwierdza jakość klasyfikacji uzyskaną przy pomocy parametru dokładności, jednocześnie nie wykazała żadnej wyróżniającej się klasy pod względem poprawności predykcji.

Badania wykazały na wystąpienie korelacji liniowej między wydźwiękiem emocjonalnym utworów a trzema cechami statystycznymi melodii - *energią*, *walencyjnością* oraz *głośnością*. Pierwsza z nich była dodatnia dla *pobudzenia* co oznacza, że wzrost *energii* w utworze idzie w parze ze wzrostem tegoż parametru. W przypadku *walencyjności* sytuacja jest odwrotna. Taka zależność oznaczać może, że utwory o wysokiej wartości *energii* w dużej mierze będą zaklasyfikowane jako *agresywne*. Analogicznie dla *głośności*. W przypadku *walencyjności* - jako cechy melodii - występuje korelacja dodatnia zarówno z *walencyjnością* jak i *pobudzeniem*. Wzrost *walencyjności* melodii idzie więc w parze ze wzrostem obydwu tych parametrów.

Zgodnie z przewidywaniami, melodie utworów *agresywnych* oraz *szczęśliwych*, w większości cechują się wysoką energią (0,6 – 1 j.m.). Oznacza to, że twórcy często łączą tę cechę z tekstami piosenek, o wyżej wskazanych wydźwiękach. Wysoka energetyczność cechuje melodie żywe, mające na celu pobudzić słuchacza. Melodie piosenek zaklasyfikowanych jako *smutne* oraz *relaksujące* nie wykazują takich tendencji. Liczba tego typu utworów rozłożona jest względnie równomiernie na całym zakresie wartości *energii*. Może to świadczyć o tym, że autorzy nie zwracają, w ich przypadku, uwagi na tę cechę lub nie wpływa ona na to czy dany utwór zostanie odebrany jako *relaksujący* lub *smutny*.

Ciekawym parametrem była walencyjność, która zgodnie z dokumentacją serwisu *Spotify*, powinna być bezpośrednio powiązana z emocjami wywoływanymi przez melodię (miara pozytywności). Dla piosenek o *szczęśliwym* tekście walencyjność najczęściej ma wartość od 0,6 do 1 j.m. Oznacza to, że zgodnie z przewidywaniem, chociaż nie aż tak jednoznacznie, pozytywność melodii idzie w parze z pozytywnym (*szczęśliwym*) wydźwiękiem tekstu utworu. W przypadku reszty piosenek (*smutnych*, *agresywnych* oraz *relaksujących*) walencyjność przyjmuje najczęściej wartości z zakresu (0,0 - 0,5 j.m.). O ile dla dwóch pierwszych wydźwięków nie jest to dziwne - niska miara tego parametru charakteryzuje melodie brzmiące negatywnie - o tyle dla trzeciego z nich takie wyniki są zaskakujące. Oznacza to, że niewielka walencyjność melodii wykorzystywana jest także, w utworach o *relaksujących* tekstach.

W przypadku głośności linii melodycznej nie było zaskoczenia. Zgodnie z przewidywaniem, głośna melodia charakteryzuje utwory o *agresywnym* oraz *szczęśliwym* tekście.

Autorzy wykorzystują tę cechę aby pogłębić emocje zawarte w słowach. Głośność muzyki pobudza, ożywia utwór. *Agresywne* oraz *szczęśliwe* teksty utworów są też często spolaryzowane. Ciche melodie charakteryzują utwory mające *relaksujący* lub *smutny* tekst. Tego typu parametry wyciszają, wywołują skupienie i zwiększają przejęcie treścią słów.

Badanie zależności ilości utworów, o określonym wydźwięku emocjonalnym tekstu, od wartości *tempa* potwierdziło brak korelacji między tymi parametrami. Oznacza to, że *tempo* melodii stosowane jest niezależnie od emocji zawartych w słowach utworów.

### 6.5. Wykorzystane narzędzia oraz współczynniki

Poniżej przedstawiono krótkie opisy wykorzystanych, w pracy magisterskiej, narzędzi oraz bibliotek.

#### 6.5.1. NLTK

*NLTK (Natural Language Toolkit)* jest wiodącą platformą zawierającą szereg bibliotek i programów - napisanych w języku *Python* - wykorzystywanych do symbolicznego i statystycznego przetwarzania języka naturalnego [38]. W jej skład wchodzi m.in. korpusy językowe, zestawy tzw. *stop words* oraz szeroki wachlarz narzędzi do obróbki tekstów. W projekcie magisterskim *NLTK* wykorzystano kilkakrotnie.

*NLTK* ma bardzo wielu oddanych fanów, jest prężnie rozwijana a w sieci dostępne jest mnóstwo samouczków oraz książek przybliżających funkcjonalności tego narzędzia.

#### 6.5.2. Sklearn

*Sklearn* jest darmową biblioteką do języka *Python* oferującą rozwiązania z zakresu uczenia maszynowego [39]. Ma na celu zapewnienie prostych i wydajnych narzędzi, które mogą być wykorzystane przez wszystkich i w różnych kontekstach.

*Sklearn* oferuje szereg rozwiązań z zakresu klasyfikacji, regresji, klasteryzacji oraz preprocessingu, umożliwia ich personalizację i zapewnia szeroką dokumentację, ułatwiającą w dużym stopniu pracę. Dodatkowo pakiet zawiera też narzędzia ewaluacyjne wykorzystywane go oceny zbudowanych modeli uczenia maszynowego.

#### 6.5.3. PyLyrics

Prosta biblioteka do języka *Python* ułatwiająca komunikację z *API* serwisu *Lyric Wikia* [36]. Oprócz wykorzystanej w pracy funkcjonalności - pobierania tekstów utworów - pakiet pozwala również na uzyskiwanie informacji o albumach oraz artystach.

Biblioteka *PyLyrics* ma otwartą licencję, umożliwiającą korzystanie z niej zarówno do celów badawczych jak i komercyjnych.

#### 6.5.4. SciPy

*SciPy* jest darmową, *pythonową* biblioteką o otwartym kodzie wykorzystywaną do obliczeń naukowych oraz technicznych [40]. Pakiet zawiera moduły do optymalizacji, algebry liniowej, całkowania, interpolacji, transformaty *Fourier'a* itp. Biblioteka opiera się na tablicach *NumPy* co czyni ją częścią pakietu *NumPy*. Obecnie *SciPy* jest rozpowszechniana na licencji *BSD* a jej rozwój jest sponsorowany przez otwartą społeczność programistów i pasjonatów.

#### 6.5.5. Współczynnik Pearsona

Współczynnik określający poziom zależności liniowej między zmiennymi losowymi [41]. Przyjmuje wartości z przedziału  $[-1, 1]$ , z czym im wartość jest bliższa 1 tym zależność jest silniejsza dodatnio, a im bliżej  $-1$  tym silniejsza ujemnie. W przypadku wartości zbliżonych do 0 nie występuje korelacja między zmiennymi. Współczynnik *Pearsona* oblicza się ze wzoru:

$$r = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2} \sqrt{\sum_i (y_i - \bar{y})^2}}$$

#### 6.5.6. Współczynnik Spearmana

Współczynnik określający poziomo dowolnej monotonicznej zależności między zmiennymi losowymi [42]. Współczynnik *Spearmana* jest uogólnioną wersją współczynnika *Pearsona*.

Współczynnik *Spearmana* bazuje na rangach, określających pozycję na której znajduje się dana obserwacja po uszeregowaniu rosnąco wszystkich obserwacji. Dla najmniejszej wartości ranga przyjmuje 1 natomiast dla największej  $n$  gdzie  $n$  to liczba obserwacji.

Podobnie jak w przypadku współczynnika *Pearsona*, przyjmuje wartości z przedziału  $[-1, 1]$  z czego im wartość jest bliższa 1 tym zależność jest silniejsza dodatnio, a im bliżej  $-1$  tym silniejsza ujemnie. Współczynnik *Spearmana* dany jest wzorem:

$$r = 1 - \frac{6 * \sum_i d_i^2}{n(n^2 - 1)}$$

gdzie  $d_i$  określa różnice między rangami  $X$  i  $Y$  natomiast  $n$  jest liczbą obserwacji.

## 7. Zakończenie

W pracy przedstawiono analizę korelacji między wydźwiękiem emocjonalnym tekstów utworów muzycznych a cechami statystycznymi melodii. Klasyfikacja wydźwięku została przeprowadzona przy pomocy metody słownikowej - wykorzystanie leksykonu *NRC* - co pozwoliło na pełną niezależność od melodii. Uzyskano klasyfikator o stosunkowo wysokiej dokładności predykcji - 57.32%. Cechy statystyczne pozyskano z otwartego *API* serwisu *Spotify*.

Badania wykazały korelację liniową między wydźwiękiem tekstów utworów a *głośnością*, *energiją* oraz *walencyjnością*. Dla skorelowanych cech zbadano zależność ilości utworów od ich natężenia. Wskazano w ten sposób powiązanie między natężeniem poszczególnych cech a wydźwiękiem emocjonalnym tekstów.

Poza badaniami w pracy przedstawiono także zestawienie metoda analizy wydźwięku tekstów, wyzwania jakie ze sobą niesie, opisano dokładnie proces klasyfikacji utworów muzycznych z wyszczególnieniem wykorzystujących nadzorowane uczenie maszynowe. Dodatkowo przedstawiono statystyczny opis melodii, sposób jego otrzymywania w praktyce oraz metodą klasyfikacji utworów przy jego pomocy.

Podczas wykonywania pracy natknięto się na kilka przeszkód, których pokonanie wymagało pewnych założeń, mogących wpłynąć na jakość analiz. Przykładowo założono, że pobierane teksty utworów były zawsze poprawne i odnosiły się do wskazanej piosenki. Do oceny zbudowanego klasyfikatora wykorzystano zbiór tagów, umieszczonych przez użytkowników *last.fm*, który ętykietyzowałały utwór a nie jedynie jego tekst. Sam sposób wyznaczania etykiet można uznać za niepewny.

W trakcie wykonywania pracy wykorzystano szereg narzędzi oraz bibliotek, które w znaczny sposób przyspieszyły pracę, poprawiły jakoś klasyfikacji oraz zapewniły niezbędną optymalizację obliczeń. Pakiety takie jak *NLTK* oraz *SciPy* znakomicie współgrają ze sobą i wzajemnie się uzupełniają. Ich olbrzymie możliwości pozwalają na bardzo szerokie zastosowanie, jednakże aby sprawnie się nimi posługiwać konieczne jest dokładne studiowanie dokumentacji, która najczęściej jest dokładna i napisana przystępnym językiem.

Kolejnym etapem badań mogło by być wykorzystanie nienadzorowanego uczenia maszynowego np. klasteryzacji danych statystycznych i próba znalezienia powiązania między stworzonymi klasami a wydźwiękiem emocjonalnych utworów, których te dane dotyczą.

Ciekawym zagadnieniem analizy było by porównanie wydźwięku utworów wokalnie instrumentalnych z wersją instrumentalną. Umożliwiło by to określenie wpływu tekstu na całościowy odbiór piosenki. W trakcie wykonywania tej pracy podjęta została próba przeprowadzenia takiej analizy, jednakże ze względu na małą liczbę utworów dostępnych w obydwu wersjach, badania zaniechano.

W celu poszerzenia badań można by przeprowadzić klasyfikację przy pomocy innej metody słownikowej - aby zachować niezależność od melodii - np. wykorzystującej reguły gramatyczne. Taki klasyfikator mógłby osiągnąć większą dokładność i być lepszym źródłem danych do badań.

W pracy wykorzystano około 10 tysięcy utworów, ale dostępnych jest ich znacznie więcej, dlatego można by spróbować powiększyć zbiór danych i sprawdzić jak jego wielkość wpływa na uzyskiwane rezultaty.

Inną metodą badań mogła by być próba powiązania cech statystycznych z gatunkami muzyki oraz ich wydźwiękiem. Obecnie klasyfikacja utworów ze względu na gatunek jest częstym tematem publikacji [43].

Cechy statystyczne, także są miejscem, w którym jest jeszcze wiele do zbadania. Możliwe jest wykorzystanie innych niż te udostępnione przez *Spotify* lub ich własnoręczne uzyskanie.

Praca spełnia założone cele - wykazuje korelację trzech cech *energii*, *głośności* oraz *walencyjności* z wydźwiękiem tekstu utworów. Wskazuje jak powyższe cechy mają się do *walencyjności* oraz *pobudzenia* tekstu. Dodatkowo praca zawiera analizę wpływu natężenia poszczególnych cech statystycznych na częstość występowania danego wydźwięku emocjonalnego.

# Bibliografia

- [1] G. Buccellati, *Hurrian music*, Dostęp zdalny (10.02.2020): <http://urkesh.org/urkeshpublic/music.htm>, 2020.
- [2] K. R. Scherer i M. R. Zentner, „Emotional effects of music: production rules”, w *Music and emotion: theory and research*. Oxford University Press, 2010, s. 361–387.
- [3] A. Gabrielle i E. Stromboli, „The influence of musical structure on emotional expression”, w *Music and emotion: theory and research*. Oxford University Press, 2010, s. 223–243.
- [4] J. X. Hu, S. Downie i A. F. Ehmann, „Lyric text mining in music mood classification”, w *International Music Information Retrieval Systems Evaluation Laboratory*, University of Illinois at Urbana-Champaign, 2009.
- [5] T. Bhavika, *Music Mood Classification Using The Million Song Dataset*, Dostęp zdalny (10.02.2020): <https://github.com/bhavika/JoyDivision/blob/master/report/report.pdf>, 2016.
- [6] J. X. Hu, S. Downie i A. F. Ehmann, „Multimodal Music Mood Classification Using Audio and Lyrics”, w *2008 Seventh International Conference on Machine Learning and Applications*, IEEE, 2008.
- [7] K. Tomanek, „Analiza sentymentu - metoda analizy danych jakościowych : przykład zastosowania oraz ewaluacja słownika RID i metody klasyfikacji Bayesa w analizie danych jakościowych”, w *Przegląd Socjologii Jakościowej 10/2*, Muzeum Historii Polski, 2014.
- [8] B. Pang i L. Lee, „Opinion mining and sentiment analysis”, w *Foundations and Trends in Information Retrieval*, Boston - Delft, 2008, s. 1–135.
- [9] B. Liu, „Sentiment Analysis and Subjectivity”, w *Handbook of Natural Language Processing*, Chapman & Hall/CRC, 2010.
- [10] X. Hu, „Music and Mood: Where Theory and Reality Meet”, lut. 2010.
- [11] C. Diamantini, A. Mircoli i D. Potena, „A Negation Handling Technique for Sentiment Analysis”, paź. 2016, s. 188–195. DOI: 10.1109/CTS.2016.0048.
- [12] L. Weitzel, R. Aguiar, P. Quaresma, T. Goncalves i R. Prati, „How Does Irony Affect Sentiment Analysis Tools?”, wrz. 2015, s. 803–808.
- [13] H. Thakkar i D. Patel, „Approaches for Sentiment Analysis on Twitter: A State-of-Art study”, grud. 2015.
- [14] A. Z. Syed, M. Aslam i A. M. Martinez-Enriquez, „Lexicon Based Sentiment Analysis of Urdu Text Using SentiUnits”, w *Advances in Artificial Intelligence*, G. Sidorov, A. Hernández Aguirre i C. A. Reyes García, red., Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, s. 32–43.
- [15] H. Zou, X. Tang, B. Xie i B. Liu, „Sentiment Classification Using Machine Learning Techniques with Syntax Features”, grud. 2015, s. 175–179. DOI: 10.1109/CSCI.2015.44.



- [16] J. Opałka, W. Abramowicz, W. Sokołowska i T. Hossa., „Automatyczna analiza wydziwku opinii o operatorach energetycznych jako element wsparcia podejmowanych decyzji”, w. Wydawnictwo Uniwersytetu Ekonomicznego w Katowicach, 2015, s. 257–273.
- [17] E. Çano, „Music Mood Dataset Creation Based on Last FM Tags”, maj 2017, s. 15–26.
- [18] X. Hu, J. Downie i A. Ehmann, „Lyric Text Mining in Music Mood Classification.”, sty. 2009, s. 411–416.
- [19] C. Gascoigne, J. Workman i Y. Zhang, *Lyric Mood Classification*, Dostęp zdalny (10.02.2020): <https://github.com/workmanjack/lyric-mood-classification/blob/master/report/lyric-mood-classification-with-deep-learning.pdf>, 2019.
- [20] R. Kaplan, „A Method for Tokenizing Text”, sty. 2005.
- [21] H. Saif, M. Fernandez i H. Alani, „On Stopwords, Filtering and Data Sparsity for Sentiment Analysis of Twitter”, maj 2014.
- [22] V. Balakrishnan i L.-Y. Ethel, „Stemming and Lemmatization: A Comparison of Retrieval Performances”, *Lecture Notes on Software Engineering*, t. 2, s. 262–267, sty. 2014.
- [23] H. Liang, X. Sun, S. Yunlei i Y. Gao, „Text feature extraction based on deep learning: a review”, *EURASIP Journal on Wireless Communications and Networking*, t. 2017, grud. 2017.
- [24] V. Dwivedi i M. Shrivastava, „Beyond Word2Vec: Embedding Words and Phrases in Same Vector Space”, grud. 2017.
- [25] H. Fan i Y. Qin, „Research on Text Classification Based on Improved TF-IDF Algorithm”, sty. 2018.
- [26] Y. Zhang, R. Jin i Z.-H. Zhou, „Understanding bag-of-words model: A statistical framework”, *International Journal of Machine Learning and Cybernetics*, t. 1, s. 43–52, grud. 2010.
- [27] D. Buzic i J. Dobša, „Lyrics Classification Using Naive Bayes”, maj 2018.
- [28] J. Zhang i C. Zong, „Deep Neural Networks in Machine Translation: An Overview”, *IEEE Intelligent Systems*, t. 30, s. 16–25, wrz. 2015.
- [29] V. Nguyen, T. Anh i H. Yang, „Real-time event detection using recurrent neural network in social sensors”, *International Journal of Distributed Sensor Networks*, t. 15, s. 155 014 771 985 649, czer. 2019.
- [30] H. Song, Z. Ding, C. Guo, Z. Li i H. Xia, „Research on Combination Kernel Function of Support Vector Machine”, t. 1, sty. 2008, s. 838–841.
- [31] M. Hossin i M. Sulaiman, „A review on evaluation metrics for data classification evaluations”, *International Journal of Data Mining & Knowledge Management Process (IJDKP)*, t. 5, nr. 2, 2015.
- [32] *Elements of Music*, Dostęp zdalny (10.02.2020): <https://wmich.edu/mus-gened/mus150/Ch1-elements.pdf>, 2016.
- [33] *Spotify api*, Dostęp zdalny (10.02.2020): <https://developer.spotify.com/>.

## 7. Bibliografia

---

- [34] M. McKinney i J. Breebaart, „Features for Audio and Music Classification”, list. 2003.
- [35] *Million Song Dataset*, Dostęp zdalny (10.02.2020): <http://millionsongdataset.com/>.
- [36] *PyLyrics*, Dostęp zdalny (10.02.2020): <https://pypi.org/project/PyLyrics/>.
- [37] *Leksykon NRC*, Dostęp zdalny (10.02.2020): <https://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm>.
- [38] *NLTK*, Dostęp zdalny (10.02.2020): <https://www.nltk.org/>.
- [39] *Sklearn*, Dostęp zdalny (10.02.2020): <https://scikit-learn.org/>.
- [40] *SciPy*, Dostęp zdalny (10.02.2020): <https://www.scipy.org/>.
- [41] P. Sedgwick, „Pearson’s correlation coefficient”, *BMJ*, t. 345, e4483–e4483, lip. 2012.
- [42] *Spearman*, Dostęp zdalny (10.02.2020): <https://geographyfieldwork.com/SpearmanRank.htm>.
- [43] R. Mayer i A. Rauber, „Music Genre Classification by Ensembles of Audio and Lyrics Features.”, sty. 2011, s. 675–680.

# Spis rysunków

3.1. Wizualizacja przestrzeni walencyjno-pobudzeniowej zaproponowanej przez <i>Russella</i>	12
4.1. Wizualizacja modelu <i>CBOW</i> w <i>word2vec</i>	21
4.2. Wizualizacja modelu <i>Continuous skip-gram</i> w <i>word2vec</i>	22
6.1. Przykładowa odpowiedź <i>API</i> serwisu <i>Spotify</i> z cechami utworu	33
6.2. <i>Word cloud</i> prezentujący słowa charakteryzujące się <i>szczęśliwym</i> wydźwiękiem, według leksykonu <i>NRC</i>	34
6.3. <i>Word cloud</i> prezentujący słowa charakteryzujące się <i>relaksującym</i> wydźwiękiem, według leksykonu <i>NRC</i>	34
6.4. <i>Word cloud</i> prezentujący słowa charakteryzujące się <i>smutnym</i> wydźwiękiem, według leksykonu <i>NRC</i>	35
6.5. <i>Word cloud</i> prezentujący słowa charakteryzujące się <i>gniewnym</i> wydźwiękiem, według leksykonu <i>NRC</i>	35
6.6. Proces analizy przy pomocy metody słownikowej wykorzystującej leksykon	37
6.7. Kolejne etapy obróbki danych	38
6.8. Wycinek wyniku algorytmu ważenia częstości termów dla danych projektowych	39
6.9. Proces klasyfikacji utworów muzycznych z wykorzystaniem leksykonu	40
6.10. Zależność liczby utworów - o określonym wydźwięku - od średniej wartości energii	45
6.11. Zależność liczby utworów od średniej wartości energii dla piosenek kolejno <i>szczęśliwych</i> , <i>agresywnych</i> , <i>smutnych</i> i <i>relaksujących</i>	46
6.12. Zależność liczby utworów - o określonym wydźwięku - od średniej wartości walencyjności	46
6.13. Zależność liczby utworów od średniej wartości walencyjności dla piosenek kolejno <i>szczęśliwych</i> , <i>agresywnych</i> , <i>smutnych</i> i <i>relaksujących</i>	47
6.14. Zależność liczby utworów - o określonym wydźwięku - od średniej wartości głośności	48
6.15. Zależność liczby utworów od średniej wartości głośności dla piosenek kolejno <i>szczęśliwych</i> , <i>agresywnych</i> , <i>smutnych</i> i <i>relaksujących</i>	49
6.16. Zależność liczby utworów - o określonym wydźwięku - od średniej wartości tempa	50
6.17. Zależność liczby utworów od średniej wartości tempa dla piosenek kolejno <i>szczęśliwych</i> , <i>agresywnych</i> , <i>smutnych</i> i <i>relaksujących</i>	50

# Spis tabel

4.1.	Rozkład etykiet emocjonalnych w przykładowym zbiorze danych . . . . .	19
4.2.	Lista słów przed i po przeprowadzeniu <i>stemmingu</i> oraz lematyzacji . . . . .	20
4.3.	Reprezentacja zdań przy pomocy metody <i>TFIDF</i> . . . . .	23
4.4.	Reprezentacja zdań przy pomocy metody <i>bag of words</i> . . . . .	24
4.5.	Przykładowa macierz błędu, dla klasyfikatora, grupującego utwory na cztery kategorie	26
6.1.	Liczba utworów zakwalifikowanych do danej kategorii . . . . .	41
6.2.	Macierz błędu dla klasyfikatora słownikowego wykorzystującego leksykon <i>NRC</i> . . .	42
6.3.	Współczynnik korelacji <i>Pearsona</i> między <i>walencyjnością</i> a kolejnymi cechami statystycznymi melodii . . . . .	43
6.4.	Współczynnik korelacji <i>Pearsona</i> między <i>pobudzeniem</i> a kolejnymi cechami statystycznymi melodii . . . . .	43
6.5.	Współczynnik korelacji <i>Spearmana</i> między <i>walencyjnością</i> a kolejnymi cechami statystycznymi melodii . . . . .	44
6.6.	Współczynnik korelacji <i>Spearmana</i> między <i>pobudzeniem</i> a kolejnymi cechami statystycznymi melodii . . . . .	44