

Temat: Analizator emocji w muzyce

Opis:

Celem projektu jest stworzenie systemu rozpoznającego emocje przekazywane przez utwory muzyczne. Program będzie analizował pliki audio i klasyfikował pod kątem emocjonalnym (zarówno całe utwory, jak i wskazane fragmenty) w trybach Valence/Arousal oraz wykorzystujący model Russela. Plan obejmuje stworzenie modelu opartego o głębokie uczenie, trenowanego na oznaczonych emocjonalnie utworach, przygotowanie danych, eksperymenty w celu poprawy jego skuteczności oraz aplikacji webowej.

Harmonogram:

16 – 22 października 2025

- Analiza wymagań i literatury z zakresu MER.
- Analiza wybranych zbiorów danych (DEAM, emoMusic, MERGE) i przygotowanie środowiska.

23 – 29 października 2025

- Implementacja prototypu bazowego na danych pozbawionych dodatkowego przeprocesowania (czyszczenia i augmentacji). Prototyp umożliwiać będzie wczytanie pliku audio, jego analizę i zwrócenie predykcji (tryb VA).
- Porównanie wyników otrzymanego prototypu z modelami dostępnymi w Essentia (tryb VA).

30 października – 9 listopada 2025

- Eksperymenty z różnymi architekturami modeli oraz analiza wpływu augmentacji danych (tryb VA).
- Analiza błędów modeli oraz próby ulepszeń.

10 – 16 listopada 2025

- Kontynuacja treningu modeli i tuning hiperparametrów (tryb VA).

17 – 23 listopada 2025

- Adaptacja zbiorów danych do zadania klasyfikacji w trybie Russel 4Q.
- Eksperymenty z różnymi architekturami modeli (tryb Russel 4Q).
- Analiza błędów modeli oraz próby ulepszeń.

24 listopada 2025 – 4 grudnia 2025

- Stworzenie aplikacji webowej:
 - wczytanie pliku audio
 - predykcja emocji
 - wizualizacja wyników
- Kontynuacja treningu modeli i tuning hiperparametrów (tryb Russel 4Q).

5 grudnia – 14 grudnia 2025

- Optymalizacja wydajności systemu.
- Przygotowanie finalnych wersji modeli (tryb VA / Russel 4Q).
- Eksperymenty przeprowadzone na finalnych wersjach modeli uwzględniające:
 - analizę wpływu długości utworu
 - test generalizacji
- Zestawienie wyników otrzymanych w trybach VA i Russel 4Q.

15 – 21 grudnia 2025

- Testy użytkowe.
- Dopracowanie UX / interfejsu.
- Przygotowywanie finalnej wersji kodu.

22 grudnia 2025 – 4 stycznia 2026

- Finalizacja projekt:
 - Dokumentacja.
 - Prezentacja
 - Nagranie demo.

Plan eksperymentów:

- Porównanie otrzymanych modeli z opublikowanymi gotowymi modelami referencyjnymi.
- Badanie wpływu cech i architektury modelu na skuteczność oraz analiza różnych sposobów reprezentacji utworu. Wyniki przedstawione zostaną w formie tabeli z metrykami (w tym: dokładność, precyzja, czułość, wynik F1) oraz wykresów.
- Porównanie działania otrzymanych modeli w trybach VA i Russell 4Q.
- Badania wpływu zastosowania uczenia transferowego.
- Badania wpływu augmentacji danych oraz wrażliwości na stopień zbalansowania próbek w klasach i technik wyrównywania niezbalansowania licznosci.
- Ocena modelu na różnych długościach utworu w celu poznania jak długość wejściowego fragmentu wpływa na stabilność predykcji. Porównanie, czy emocja jest łatwo rozpoznawana niezależnie od fragmentu oraz spójna z emocją przewidzianą w kontekście całego utworu.
- Porównanie oceny emocji użytkowników dla utworów z etykietami zbioru danych i predykcjami modelu. To pomoże ocenić, na ile rozbieżności modelu wynikają z trudności samego zadania (np. ludzie też się nie zgadzają co do emocji utworu), a na ile z błędów modelu.

Planowana funkcjonalność programu:

- Interfejs użytkownika realizowany poprzez aplikację webową
- Analiza emocji z pliku audio:
 - Użytkownik może wczytać plik muzyczny poprzez interfejs.
 - Program dokonuje analizy i wyświetla wynik w postaci rozpoznanej emocji włącznie ze stopniem pewności.
 - Możliwość wyboru analizowanego fragmentu.
 - Analiza zmian emocji w czasie przy regulowanym oknie czasowym.
- Możliwość wyboru trybu pracy: Valence/Arousal oraz wykorzystujący model Russela.
- Możliwość porównania skuteczności różnych modeli w zadaniu rozpoznawaniu emocji w muzyce.
- Wizualizacja otrzymywanych predykcji
 - Predykcja uwzględniająca cały kontekst utworu.
 - Predykcja uwzględniająca kontekst zaznaczonego fragmentu utworu.
 - Prezentacja zmiany emocji w czasie.

Stack technologiczny:

- Kontrola wersji:
 - git
- Język:
 - Python
- Biblioteki:
 - pytorch (ML)
 - tensorflow (ML)
 - streamlit (UI)
 - numpy (przetwarzanie danych)
 - pandas(przetwarzanie danych)
 - torchaudio (przetwarzanie audio; przygotowanie danych treningowych),
 - librosa (przetwarzanie audio)
- Linter:
 - flake8
- Autoformatter:
 - black
- Testy automatyczne:
 - pytest (testy jednostkowe)
 - tox (testowanie na różnych wersjach interpretera)
- Środowisko wirtualne:
 - venv

Bibliografia

- Liyanarachchi, R., Joshi, A., & Meijering. "A Survey on Multimodal Music Emotion Recognition." *arXiv:2504.18799*, 2025
- Louro, P.L. i in. "A Comparison Study of Deep Learning Methodologies for Music Emotion Recognition." *Sensors*, 24(7):2201, 2024.
- Kang, J., Herremans, D. "Are we there yet? A brief survey of Music Emotion Prediction Datasets, Models and Outstanding Challenges." *arXiv:2406.08809*, 2024
- Louro P. L., Redinho H., Ribeiro T. T. F., Santos R, Malheiro R., Panda R. & Paiva R. P. (2025). "MERGE - A Bimodal Audio-Lyrics Dataset for Static Music Emotion Recognition".preprint. [Online]. Available: <http://arxiv.org/abs/2407.06060>.
- <https://essentia.upf.edu/models.html>
- <http://mir.dei.uc.pt/downloads.html>
- <https://cvml.unige.ch/databases/DEAM/>
- <https://cvml.unige.ch/databases/emoMusic/>