



**Politechnika
Śląska**

**Wydział Automatyki, Elektroniki
i Informatyki**

Systemy Interaktywne i Multimedialne
Projekt
Detekcja emocji w głosie

Natalia Stręk, Jakub Kula, Paweł Wójtowicz

Gliwice 2023

Spis treści

1	Wizja projektu	2
2	Cele projektu	2
3	Zakres projektu	2
4	Etapy w projekcie	3
5	Charakterystykę narzędzi	3

1 Wizja projektu

Projekt ten ma na celu stworzenie systemu zdolnego do detekcji emocji w ludzkim głosie takich jak radość, smutek, złość czy strach. Ponieważ ludzie głównie rozpoznają emocje poprzez komunikację niewerbalną, efektywna sztuczna inteligencja powinna więc być zdolna do lepszego rozpoznawania emocji w głosie człowieka, wykorzystując do tego nieoczywiste i niezauważalne wzorce, które są trudne do dostrzeżenia na pierwszy rzut oka. System będzie zrealizowany przy wykorzystaniu technologii do przetwarzania sygnałów dźwiękowych jak i sztucznej inteligencji. Realizacja tego projektu może przynieść korzyści w usprawnieniu interakcji człowiek-maszyna dzięki czemu maszyna będzie mogła lepiej zrozumieć intencje oraz potrzeby użytkownika. Inną korzyścią płynącą z tego projektu może być badanie nad dynamiką emocji w mediach społecznościowych i transmisjach na żywo co pozwoli na lepsze zrozumienie reakcji społecznych. Wraz z rozwojem projektu, możliwe jest również zastosowanie detekcji emocji w ludzkim głosie obszarach takich jak medycyna. Takie systemy mogą być szczególnie pomocne w diagnozowaniu i leczeniu zaburzeń zdrowia psychicznego.

2 Cele projektu

- Opracowanie modelu uczenia maszynowego, który będzie potrafił rozpoznawać emocje w nagraniach głosowych.
- Stworzenie interaktywnego narzędzia umożliwiającego przesłanie nagrania głosowego i otrzymanie informacji zwrotnej na temat emocji wykrytych w głosie.

3 Zakres projektu

Projekt obejmuje stworzenie systemu zdolnego do detekcji szerokiego zakresu emocji w ludzkim głosie, włączając w to radość, smutek, złość, zaskoczenie, obojętność, spokój, odraza i strach. Wykorzystując technologie do przetwarzania sygnałów dźwiękowych oraz sztucznej inteligencji, projekt będzie skoncentrowany na identyfikację wzorców w ludzkim głosie, które mogą wskazywać na określoną emocję. Kluczowym etapem projektu będzie odpowiednie przygotowanie danych polegające na ekstrakcję cech z sygnałów dźwiękowych a następnie odpowiednim przetworzeniu.

W ramach projektu zostanie opracowany i przetestowany model uczenia maszynowego, który zostanie wytrenowany na zestawie danych zawierających nagrania w języku angielskim. Projekt zakłada stworzenie modelu który będzie osiągał określoną dokładność w rozpoznawaniu emocji w ludzkim głosie. Ponadto przewiduje się stworzenie interaktywnego interfejsu użytkownika umożliwiającego przesłanie nagrania głosowego oraz uzyskanie odpowiedzi na tematy wykrytych emocji.

Projekt skupia się wyłącznie na analizie emocji w ludzkim głosie i nie obejmuje dodatkowych analiz, takich jak funkcje przetwarzania języka naturalnego w celu lepszego zrozumienia kontekstu emocjonalnego, ani rozpoznawanie emocji w innych formach komunikacji niewerbalnej, takich jak mimika twarzy czy wykonywane gesty. W związku z faktem, że model uczenia maszynowego zostanie trenowany na zbiorze danych zawierających nagrania w języku angielskim, istnieje ryzyko wystąpienia trudności w identyfikacji emocji pochodzących z nagrań w innych językach. W konsekwencji projekt nie będzie obejmował klasyfikacji emocji związanych z nagraniami w innych języku niż angielskim. Dodatkowo, projekt skoncentrowany jest wyłącznie

na identyfikacji emocji w nagraniach głosowych i nie będzie obejmować analizy innych źródeł danych.

4 Etapy w projekcie

- Zgromadzenie i przetworzenie danych - dane w postaci krótkich nagrań dźwiękowych reprezentujące różne emocje.
- Wybór oraz ekstrakcja cech.
- Wybór architektury sieci, hiperparametrów, podział danych na zbiór uczący oraz trenin-gowy.
- Trenowanie modelu i optymalizacja parametrów modelu.
- Ocena i walidacja modelu.
- Dostosowywanie parametrów modelu w celu poprawy jego wydajności.
- Stworzenie UI pozwalającego użytkownikowi dodanie pliku dźwiękowego oraz nagrania własnej wiadomości przy pomocy wbudowanego mikrofonu
- Testowanie projektu

5 Charakterystykę narzędzi

- Python - główny język programowania. W nim zostaną napisane skrypty przetwarzające dane oraz zostanie stworzony i nauczony model sieci neuronowe.
- Tensorflow i keras - Tensorflow będąca najpopularniejszą biblioteką uczenia maszynowego stanowi główne narzędzie do tworzenia sieci neuronowych. Keras udostępnia API do tworzenia modeli uczenia maszynowego.
- Numpy i Pandas - fundamentalne biblioteki do obliczeń naukowych w pythonie, umożliwiające efektywne przetwarzanie dużych ilości danych oraz ich wnikliwą analizę
- Librosa - biblioteka pozwalająca w łatwy sposób otworzyć i przetworzyć dane audio.
- Skit-lern, Eli5 - biblioteki zawierające między innymi bogaty zbiór algorytmów uczenia maszynowego. Pozwalające na uproszczenie etapu tworzenia sieci.
- Conda - Conda jest środowiskiem wirtualnym i systemem do zarządzania pakietami. Pozwala ona na tworzenie odseparowanych środowisk dla różnych projektów.
- Tkinter - biblioteka pozwalająca nam uprościć proces tworzenia działającego interfejsu graficznego.