

Temat projektu:

**Analiza gestów twarzy z wykorzystaniem
Grammatical Facial Expressions**

Autorzy:

Zuzanna Jaroszczyk (s27766)

Vladyslav Rudko (s28611)

Data: 20.01.2025

Contents

Cel badania.....	3
Opis zbioru danych	3
Rozkład klas decyzyjnych:.....	4
Wnioski z rozkładu klas	5
Metodologia i rozwiązanie	5
Wstępne przetwarzanie danych.....	5
Budowa i optymalizacja modeli.....	6
Ocena jakości modeli	6
Wyniki eksperymentalne	10
Podsumowanie i wnioski	13
Komentarze i analiza wyników.....	13
Wnioski.....	13

Cel badania

Celem badania jest opracowanie modelu klasyfikacyjnego zdolnego do rozpoznawania **Gramatycznych Wyrażeń Twarzy** (Grammatical Facial Expressions - GFE) w kontekście języka migowego używanego w Brazylii (Libras). Wyrażenia te odgrywają kluczową rolę w przekazywaniu informacji gramatycznych i semantycznych w języku migowym, zastępując intonację występującą w językach mówionych.

Automatyczne rozpoznawanie tych wyrażeń umożliwi tworzenie systemów wspomagających komunikację, takich jak tłumacze języka migowego lub systemy analizy emocji i gestów.

Znaczenie badania

Rozpoznawanie wyrażeń twarzy jest istotnym zagadnieniem w różnych dziedzinach, takich jak biometria, analiza emocji i języki wizualne. Dzięki zastosowaniu technologii takich jak Microsoft Kinect i metod uczenia maszynowego, badanie umożliwia bardziej precyzyjne przetwarzanie i interpretację gestów w Libras.

Opis zbioru danych

Źródło danych

Zbiór danych **Grammatical Facial Expressions** został opracowany przez zespół badawczy z **Uniwersytetu São Paulo** oraz firmę **Incluir Tecnologia LTDA ME** w Brazylii. Zawiera nagrania wideo użytkowników wykonujących różne gramatyczne wyrażenia twarzy związane z Libras.

- **Łączna liczba próbek:** 27,936 klatek.
- **Liczba cech:** 301 (współrzędne x, y, z dla różnych punktów twarzy).

Struktura danych

Dane są zorganizowane w 36 plikach:

- 18 plików z punktami twarzy (*_datapoints.txt).
- 18 plików z etykietami klas (*_targets.txt).

Przykładowe punkty twarzy:

- Oczy, brwi, nos, usta, kontur twarzy, źrenice.

Rozkład klas

Zbiór danych obejmuje 18 różnych wyrażień twarzy, w tym:

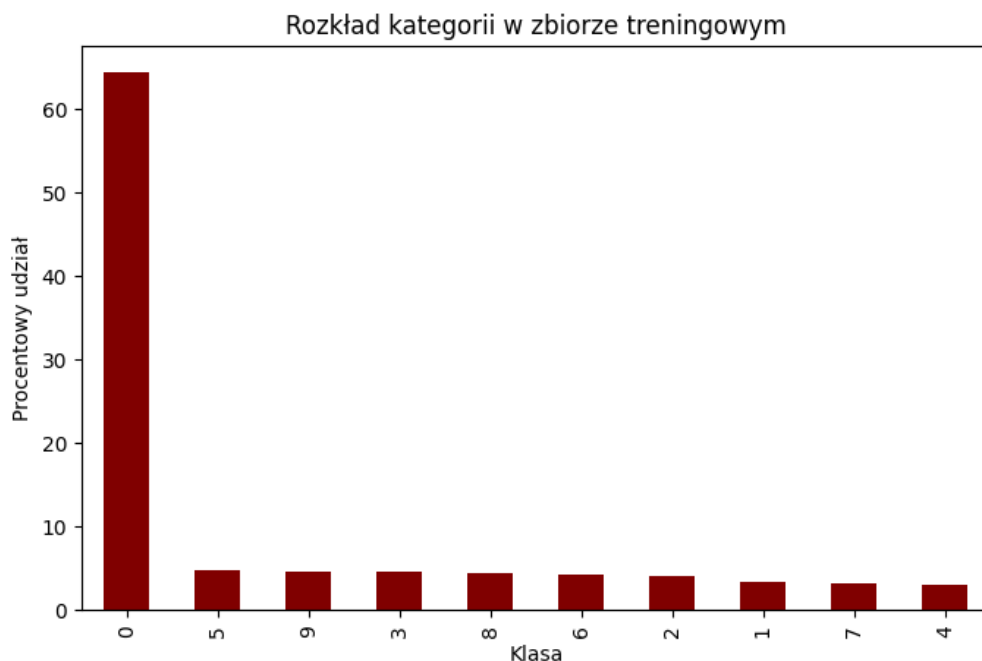
- **affirmative, conditional, doubt, emphasis, negative, relative, topics, wh_question, yn_question.**
- Dodatkowo klasa "unknown expression" (oznaczająca brak gramatycznego wyrażenia).

```
Liczba rekordów w zbiorze treningowym: 22348  
Liczba cech w zbiorze treningowym: 301  
Liczba rekordów w zbiorze testowym: 5588  
Liczba cech w zbiorze testowym: 301
```

Rozkład klas decyzyjnych:

Zbiór danych składa się z **10 różnych klas decyzyjnych**, reprezentujących różne gesty twarzy:

- Klasy: ['affirmative', 'conditional', 'doubt', 'emphasis', 'negative', 'relative', 'topics', 'wh_question', 'yn_question', 'unknown expression'].
- Dane są niezbalansowane: klasa **0 (unknown expression)** stanowi dominującą większość, z udziałem ponad 64% w obu zbiorach.
- Pozostałe klasy mają udział na poziomie około 2-5% każda.



Wnioski z rozkładu klas:

- Klasa **'unknown expression'** dominuje, co wskazuje na niezbalansowany charakter zbioru danych.
- Klasa **0 (unknown expression)** dominuje w obu zbiorach, co sugeruje większy wpływ tej klasy na uczenie modeli.
- Rzadziej występujące klasy, takie jak **'wh_question'** czy **'relative'**, mogą stanowić wyzwanie dla modeli klasyfikacyjnych.

Metodologia i rozwiązanie

Wstępne przetwarzanie danych

Przed przystąpieniem do budowy modeli, dane zostały odpowiednio przygotowane i przetworzone. Proces przetwarzania obejmował następujące kroki:

1. Scalanie danych:

- Zebrano dane z 36 plików (18 plików z punktami - datapoints oraz 18 plików z etykietami - targets) i połączono je w jedną całość.
- Współrzędne przestrzenne (x, y, z) oraz etykiety klas zostały odpowiednio dopasowane.

2. Normalizacja danych:

- Wszystkie cechy zostały znormalizowane, aby każda z nich miała wartość w zakresie [0, 1]. To kluczowe dla efektywności sieci neuronowych.

3. Podział danych:

- Zbiór danych został podzielony na:
 - **Zbiór treningowy:** 80% danych.
 - **Zbiór testowy:** 20% danych.
- Łączna liczba rekordów:
 - **Zbiór treningowy:** 22,348 rekordów.
 - **Zbiór testowy:** 5,588 rekordów.

4. Obsługa niezbalansowanego zbioru:

- W danych klasa "unknown expression" stanowiła dominującą większość (około 64% próbek). Aby poprawić jakość modeli, zastosowano wagę klas (class_weight='balanced') w modelach RandomForest oraz dopracowano architekturę sieci neuronowych.

Budowa i optymalizacja modeli

1. RandomForest:

- Przeprowadzono eksperymenty z różnymi hiperparametrami:
 - `n_estimators`: 50, 100, 150, 200.
 - `max_depth`: 10, 20, 30, 40.
- Najlepsze hiperparametry:
 - **`n_estimators=150, max_depth=30`**.
 - **Accuracy: 0.9352**.
- Analizowano wyniki za pomocą:
 - Heatmapy zależności między `n_estimators` a `max_depth`.
 - Wykresów przedstawiających wpływ hiperparametrów na dokładność.

2. Sieci neuronowe:

- **Model Variation 1:**
 - Architektura z 2 warstwami ukrytymi:
 - **Liczba neuronów:** 3x liczba cech oraz 2x liczba cech.
 - Aktywacja: ReLU.
 - Wyjście: Softmax (klasyfikacja wieloklasowa).
- **Model Variation 2:**
 - Architektura z 6 warstwami ukrytymi:
 - Liczba neuronów: 512, 256, 128, 64, 32, 16.
 - Aktywacja: ReLU.
 - Wyjście: Softmax.
- **Trening:**
 - Liczba epok zmieniana w zakresie: 10, 50, 100, 150, 200, 250.
 - Optymalizator: Adam.
 - Metryka: Accuracy.

Ocena jakości modeli

1. Metryka oceny:

- Dokładność (Accuracy) jako główna miara skuteczności.
- F1-score jako dodatkowa metryka (w przypadku RandomForest).

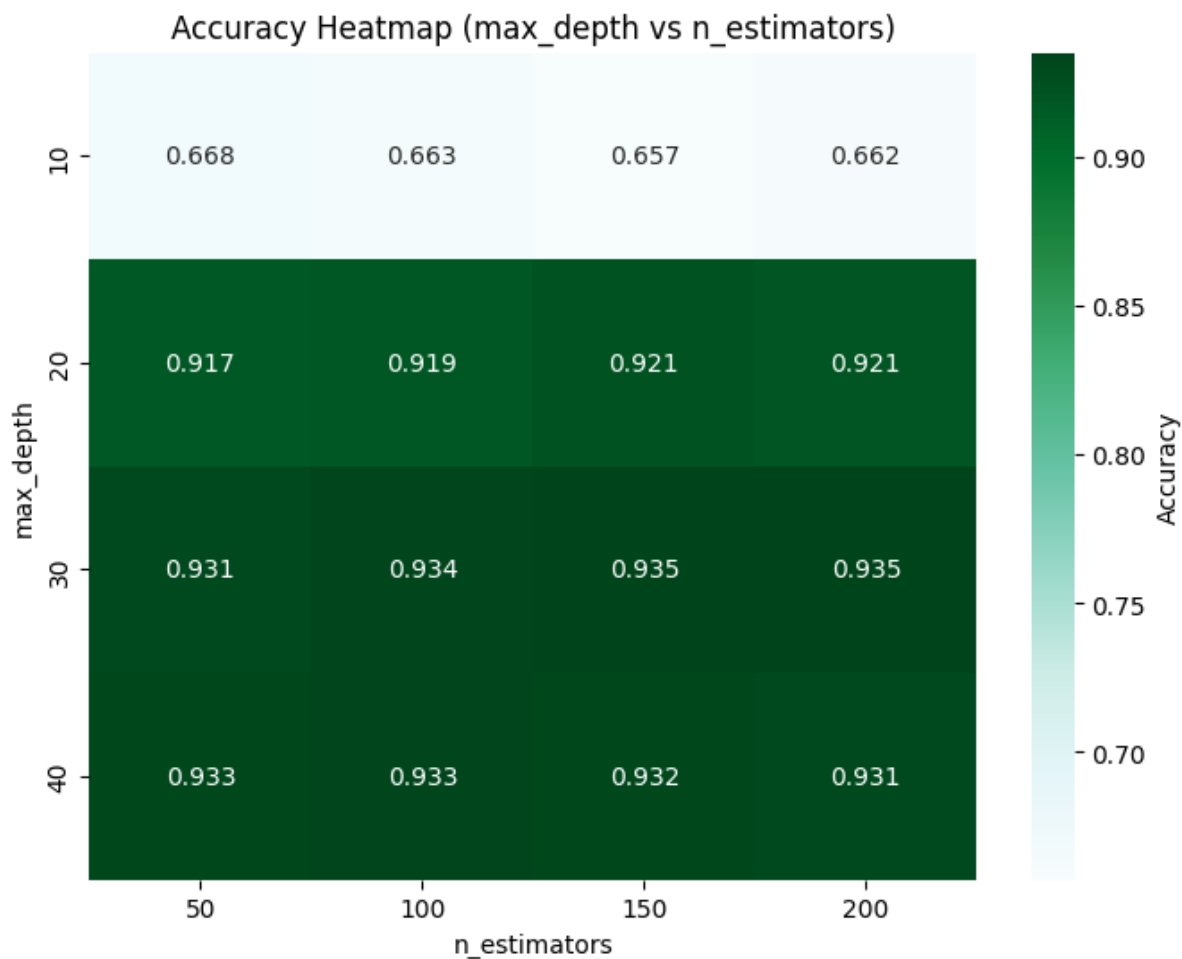
2. Macierz konfuzji:

- Obliczana dla modeli RandomForest i sieci neuronowych.
- Wskazuje poprawne oraz błędne przypisania klas, co pozwala na szczegółową analizę jakości predykcji.

Narzędzia wizualizacji wyników:

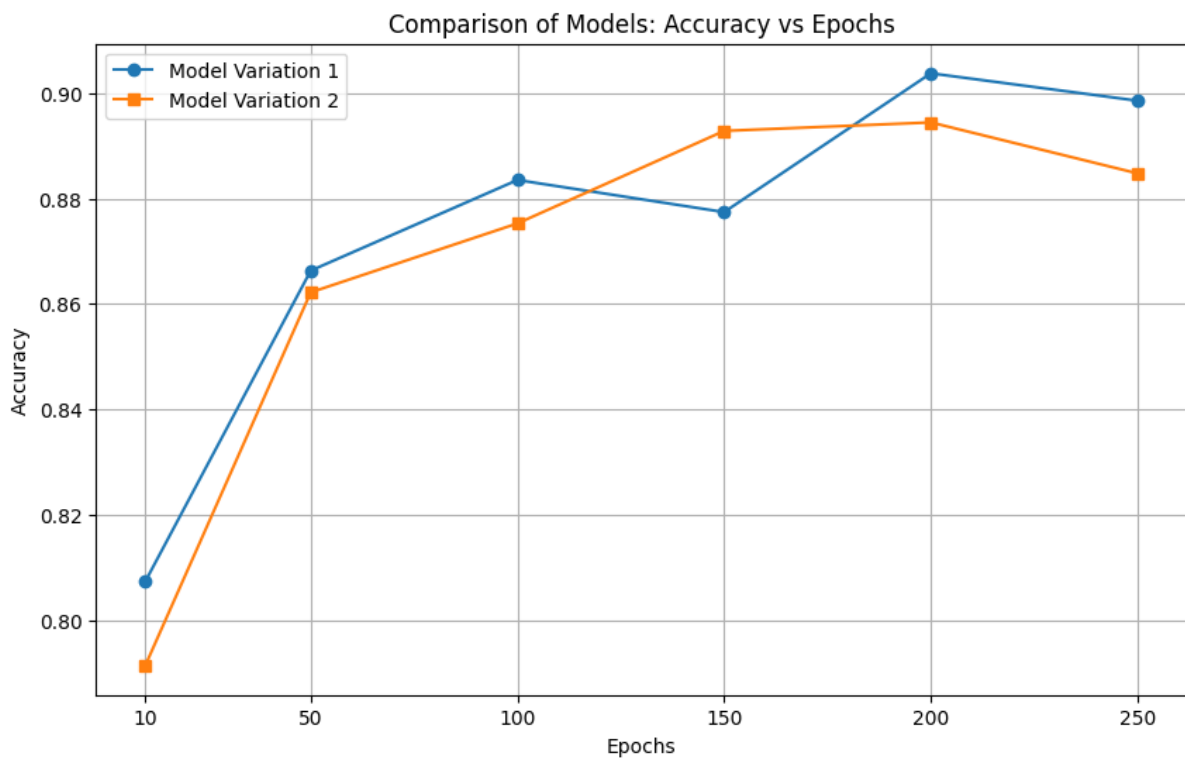
1. Heatmapa dla RandomForest:

- Pokazuje zależność dokładności od liczby drzew (`n_estimators`) i głębokości drzewa (`max_depth`).



2. Wykresy Accuracy vs Epochs:

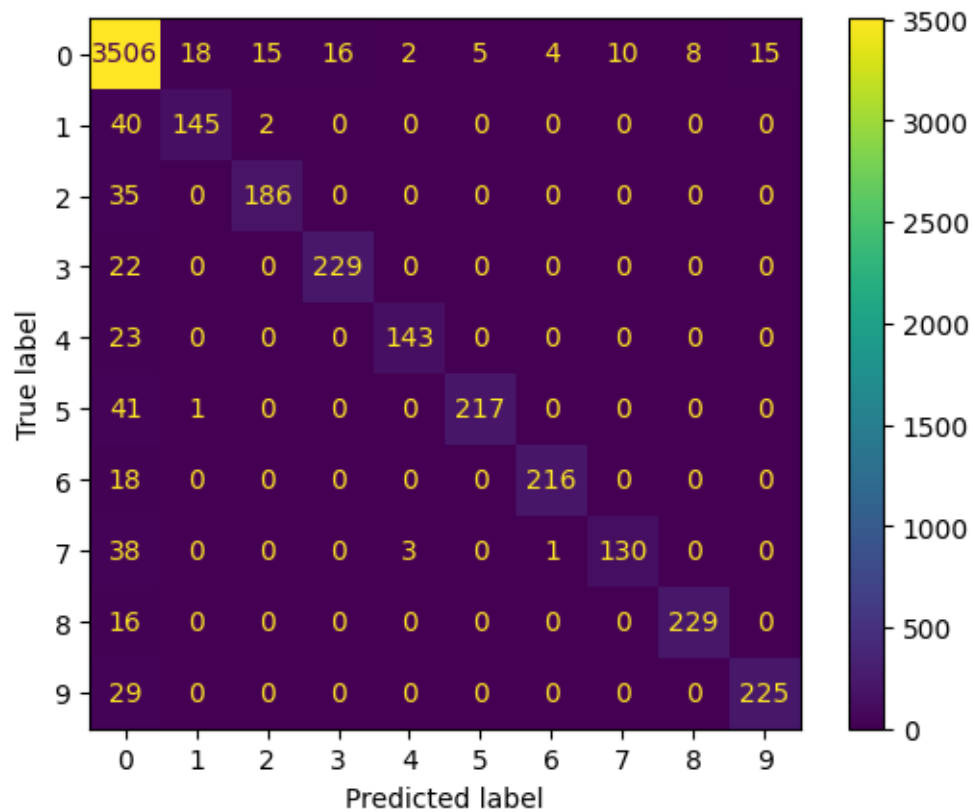
- Porównanie dwóch modeli sieci neuronowych (Variation 1 i 2) dla różnych liczby epok.



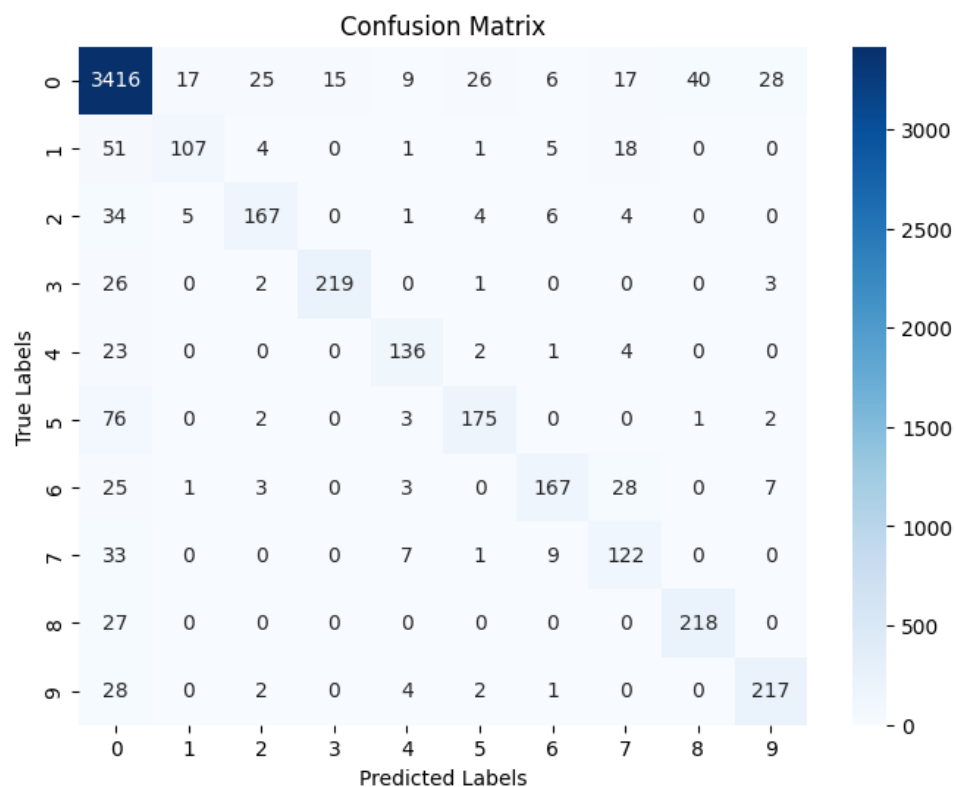
3. Macierz konfuzji:

- Wykresy prezentujące liczbę poprawnych i błędnych klasyfikacji.

Dla RandomForest:



Dla sieci neuronowej:



Wyniki eksperymentalne

1. Wyniki modelu RandomForest

Najlepsze wyniki:

- Najlepsze hiperparametry:
 - n_estimators: **150**.
 - max_depth: **30**.
- Osiągnięta dokładność (Accuracy): **93.52%**.

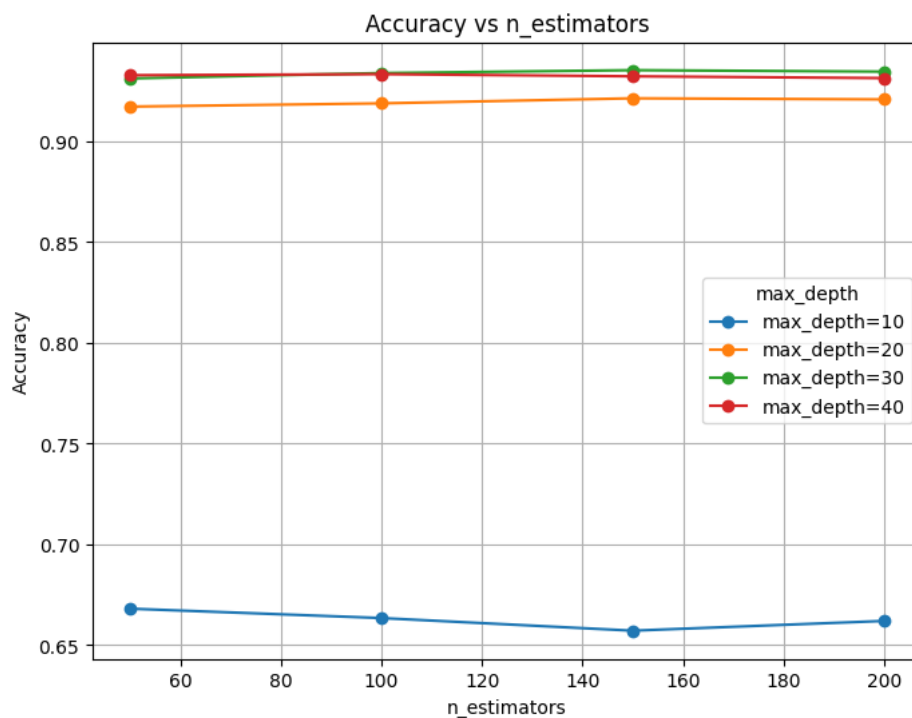
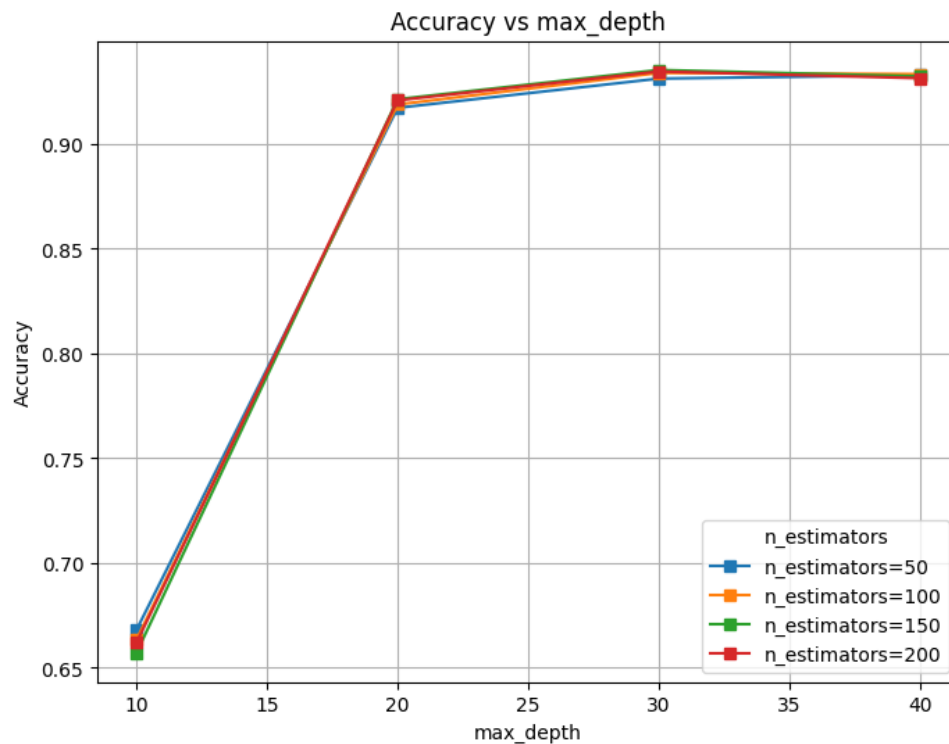
Analiza wyników:

- Wpływ hiperparametrów:
 - Głębsze drzewa (do max_depth=30) poprawiały wyniki, jednak dalsze zwiększanie głębokości nie przynosiło znaczącej poprawy.
 - Liczba drzew (n_estimators) powyżej 150 również nie wpływała znacząco na wyniki.

Wizualizacje wyników:

1. Wykresy dokładności w zależności od liczby drzew i głębokości:

- Pokazuje stabilność dokładności przy optymalnych parametrach.



2. Wyniki sieci neuronowych

Model Variation 1:

- Architektura z 2 warstwami ukrytymi.
- Maksymalna dokładność: **90.1%** (przy 200 epokach).

Model Variation 2:

- Architektura z 6 warstwami ukrytymi.
- Maksymalna dokładność: **90.7%** (przy 250 epokach).

Porównanie modeli:

- **Model Variation 2** osiągnął nieco lepsze wyniki dzięki bardziej złożonej architekturze, ale różnica była marginalna.
- **Stabilność:** Przy większej liczbie epok oba modele były stabilne, z tendencją do delikatnego spadku dokładności po 250 epokach.

3. Wnioski z wyników

- **RandomForest:**
 - Model ten osiągnął najwyższą dokładność w porównaniu do sieci neuronowych (**93.52%**).
 - Jest bardziej odporny na niebalansowany zbiór dzięki zastosowaniu wag klas.
- **Sieci neuronowe:**
 - Osiągnęły dobrą dokładność (~90%), ale wymagały więcej zasobów obliczeniowych.
 - Przy większej liczbie epok dokładność wzrastała, ale po 200 epokach modele stawały się bardziej niestabilne

Podsumowanie i wnioski

Komentarze i analiza wyników

- **RandomForest:**

- Model RandomForest osiągnął najlepsze wyniki spośród wszystkich testowanych modeli, uzyskując dokładność **93.52%**.
- Najlepsze wyniki uzyskano przy liczbie drzew `n_estimators=150` i maksymalnej głębokości `max_depth=30`.
- Model był stabilny i odporny na niebalansowany charakter zbioru danych dzięki zastosowaniu wag klas (`class_weight='balanced'`).

- **Sieci neuronowe:**

- Oba warianty sieci neuronowych (Variation 1 i Variation 2) osiągnęły dobre wyniki (~90%), ale nie przewyższyły RandomForest.
- Wzrost liczby epok poprawiał wyniki do pewnego momentu, po czym modele zaczęły tracić stabilność (przejawy przeuczenia).
- Model Variation 2 (z bardziej złożoną architekturą) osiągnął nieco wyższą dokładność, ale różnica w stosunku do prostszego modelu Variation 1 była minimalna.

- **Macierz konfuzji:**

- Klasa "unknown expression" była klasyfikowana z najwyższą skutecznością w obu modelach.
- Rzadziej występujące klasy, takie jak "relative" i "yn_question", były trudniejsze do poprawnej klasyfikacji, co jest typowe dla niebalansowanych zbiorów danych.

Wnioski

1. Najlepszy model:

- RandomForest jest najlepszym modelem dla tego zadania dzięki swojej skuteczności, stabilności oraz odporności na niebalansowany charakter zbioru danych.

2. Wyzwania i ograniczenia:

- **Niebalansowany zbiór danych:** Klasa "unknown expression" dominowała w danych, co wpłynęło na jakość predykcji innych klas.
- **Trudność w klasyfikacji rzadkich klas:** Klasy o małej liczbie próbek wymagały dodatkowych technik przetwarzania danych, takich jak oversampling, aby poprawić skuteczność klasyfikacji.

3. Praktyczne zastosowanie:

- Model RandomForest może być bezpośrednio wykorzystany do klasyfikacji gestów twarzy, np. w aplikacjach komputerowego rozpoznawania emocji lub interfejsach człowiek-maszyna.