Projekt Wczesnego Wykrywania Raka Piersi

**Wiktoria Grzesik**

06-DUMALI0 2023/SZ

# Cel projektu

Celem projektu było stworzenie modelu klasyfikacji obrazów mammografii w celu rozpoznawania potencjalnych zmian nowotworowych w tkance piersi. Klasyfikacja ta może pomóc w wczesnym wykrywaniu raka piersi.

# Dane

Dane pochodzą ze zbioru danych zawierających obrazy mammografii:

<https://archive.ics.uci.edu/dataset/17/breast+cancer+wisconsin+diagnostic>  
<https://www.kaggle.com/datasets/uciml/breast-cancer-wisconsin-data/data>

Po odrzuceniu obserwacji odstających, uzyskano 569 przykładów, które podzielono na zbiór uczący (455 przykłady) i zbiór testowy (114 przykładów).

# Modele

W projekcie porównano działanie 3 modeli:

* Regresja logistyczna
* Drzewa decyzyjne
* Sieci neuronowe

# Ewaluacja

Do ewaluacji wykorzystano metryki: **accuracy, precision, recall i F1-score**.   
Wyniki ewaluacji przedstawia poniższa tabelka:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model** | **Accuracy** | **Precision** | **Recall** | **F1 Score** |
| **Regresja logistyczna** | 96% | 96% | 94% | 95% |
| **Drzewa decyzyjne** | 91% | 91% | 87% | 89% |
| **Sieci neuronowe** | 95% | 98% | 89% | 93% |

# Wnioski

Regresja logistyczna osiągnęła najlepsze wyniki F1-score, zachowując równowagę między precyzją a odzyskiem. W przypadku, gdy priorytetem jest precyzja, model sieci neuronowej może być bardziej odpowiedni. Natomiast model drzewa decyzyjnego uzyskał najniższe wyniki we wszystkich metrykach, sugerując, że może być mniej skuteczne w danym zadaniu.