

## 1. Cel badania

Celem badania jest budowa optymalnego modelu klasyfikacyjnego za pomocą wybranych algorytmów uczenia maszynowego oraz przeprowadzenie jego walidacji. Zadaniem modelu będzie zweryfikowanie na podstawie podanych predyktorów typu guza mózgu.

## 2. Opis zbioru badawczego

Zbiór danych Brain Tumor Dataset pochodzi ze strony kaggle. Zbiór ten zawiera symulowane dane dotyczące diagnozy guza mózgu, leczenia i szczegółów pacjenta. Składa się z 19 kolumn i 20 000 wierszy, dostarczając informacji takich jak dane demograficzne pacjenta, charakterystyka guza, objawy, szczegóły leczenia i wymagania dotyczące obserwacji.

Kolumny użyte w badaniu zawierają informacje, takie jak:

- 1) Patient\_ID: Unikalny identyfikator każdego pacjenta – zmienna numeryczna.
- 2) Wiek: Wiek pacjenta (w latach) - zmienna numeryczna.
- 3) Płeć: Płeć pacjenta - zmienna binarna.
- 4) Tumor\_Type: Typ guza - zmienna binarna (Malignant/Benign), zmienna zależna.
- 5) Tumor\_Size: Rozmiar guza w centymetrach - zmienna numeryczna.
- 6) Location: Część mózgu, w której znajduje się guz – zmienna jakościowa, przyjmuje poziomy:
  - a) Temporal,
  - b) Parietal,
  - c) Frontal,
  - d) Occipital.
- 7) Histology: Typ histologiczny guza – zmienna jakościowa, przyjmuje 4 poziomy:
  - a) Astrocytoma ,
  - b) Glioblastoma,
  - c) Frontal,
  - d) Occipital.
- 8) Stadium: Stadium guza – zmienna jakościowa, przyjmuje 4 poziomy (I, II, III, IV).
- 9) Symptom\_1: Pierwszy zaobserwowany objaw - zmienna jakościowa, przyjmuje 4 poziomy:
  - a) Vision Issues,
  - b) Headache,
  - c) Seizures,
  - d) Nausea.
- 10) Symptom\_2: drugi zaobserwowany objaw - zmienna jakościowa, przyjmuje 4 poziomy:
  - a) Vision Issues,
  - b) Headache,

- c) Seizures,
- d) Nausea.

11) Symptom\_3: trzeci zaobserwowany objaw - zmienna jakościowa, przyjmuje 4 poziomy:

- a) Vision Issues,
- b) Headache,
- c) Seizures,
- d) Nausea.

12) Radiation\_Treatment: Czy zastosowano radioterapię - zmienna binarna (Yes/No).

13) Surgery\_Performed: Czy przeprowadzono operację - zmienna binarna (Yes/No).

14) Chemotherapy: Czy zastosowano chemioterapię - zmienna binarna (Yes/No).

15) Survival\_Rate: Szacowany wskaźnik przeżycia pacjenta (w procentach) – zmienna numeryczna.

16) Tumor\_Growth\_Rate: Tempo wzrostu guza (cm na miesiąc) - zmienna numeryczna.

17) Family\_History: Czy w rodzinie pacjenta występowały nowotwory mózgu - zmienna binarna (Yes/No).

18) MRI\_Result: Wynik badania MRI - zmienna binarna (Positive/Negative).

19) Follow\_Up\_Required: Czy wymagana jest obserwacja - zmienna binarna (Yes/No).