

# Raport projektu – Uczenie Maszynowe

Autor: Marek Raczkowski 481894

---

## Cel projektu:

Celem było stworzenie modelu, który przewidzi potencjalne zaangażowanie gracza w grze online (tzw. engagement level) przekładające się na to, na jak długo gracz w niej zostanie.

Zbiór zawiera wiele cech takich jak: dane demograficzne graczy, informacje o samej grze jak i zdobyte w niej osiągnięcia.

## Zbiór danych:

Dane pochodzą z serwisu kaggle.com i zostały skomponowane przez Mr. Rabie El Kharoua (Link: <https://www.kaggle.com/datasets/rabieelkharoua/predict-online-gaming-behavior-dataset?resource=download>).

Obróbkę danych przeprowadziłem w osobnym pliku o nazwie „preprocess.py” – po przetworzeniu i zbalansowaniu klas otrzymałem równo 30 000 przykładów (30000 wierszy i 22 kolumny). Po podziale na zbiory, sytuacja wygląda następująco:

- zbiór uczący – 24 000 przykładów (80%)
- zbiór testowy – 6 000 przykładów (20%)

**Uwaga:** Dla wygody, przetworzone dane zapisałem do nowego pliku o nazwie „preprocessed\_data.csv”.

## Modele:

W projekcie stworzyłem i porównałem 3 modele:

- Sieć neuronowa 3-warstwowa (4 razem z warstwą wejściową). Użyłem interfejsu *keras* z biblioteki *Tensorflow*. Jako optymalizatora wykorzystano „Adam”. Wielkość batcha wynosi: 32, a ilość epok: 100.
- Regresja logistyczna wielomianowa (1 stopnia). Zastosowałem regularyzację L2.
- Naiwny klasyfikator bayesowski. Użyłem rozkładu normalnego (Gausa) jako modelu.

**Uwaga:** projekt rozbiłem na osobne pliki – po jednym na każdy z powyższych modeli.

## Ewaluacja:

Do ewaluacji skorzystałem z metryk: *accuracy*, *precision*, *recall* i *F1-score*. Dla 3 ostatnich użyłem średniej ważonej wyników dla poszczególnych klas.

Wyniki zebrałem w poniższej tabeli:

Model	Accuracy	Precision	Recall	F1-score
Sieć neuronowa 3-warstwowa	89.37%	89.62%	89.36%	89.41%
Regresja logistyczna z regularyzacją	80.50%	80.35%	80.50%	80.40%
Naiwny klasyfikator bayesowski	80.05%	80.11%	80.05%	79.85%

Wnioski: