

# Raport z projektu

06-DUMAU10 2021/SL

## Cel projektu

Celem projektu było stworzenie modelu, który przewiduje typ domu na podstawie cen nieruchomości, powierzchni, liczby pokoi oraz lokalizacji szwajcarskich nieruchomości.

## Dane

Dane pochodzą z platformy kaggle link:

<https://www.kaggle.com/datasets/etiennekaiser/switzerland-house-price-prediction-data>

Brakujące dane zostały uzupełnione w przypadku wartości numerycznych średnią a w przypadku łańcuchów znaków najczęściej powtarzającym się znakiem/wyrazem. Uwzględniając łącznie jest 16701 przykładów w tym zbiór uczący wynosi 13361 przykładów a testowy 3340.

## Modele

W projekcie porównano działanie 4 modeli:

- Regresja logistyczna wielomianowa 3. stopnia. Jako optymalizatora użyto mini-batch gradient descent z wielkością batcha równą 100. Zastosowano regularyzację L2.
- Regresja logistyczna wielomianowa 3. stopnia. Jako optymalizatora użyto mini-batch gradient descent z wielkością batcha równą 100. Brak regularyzacji.
- Naiwny klasyfikator bayesowski. Jako modelu klas użyto rozkładu normalnego.
- Model sieci neuronowej. Jako architekturę sieci użyto dwuwarstwową sieć neuronową z 32 neuronami w warstwie ukrytej. Jako funkcję aktywacji na warstwie ukrytej użyto ReLUa na warstwie wyjściowej Softmax. Jako optymalizator został użyty algorytm Adam.

## Ewaluacja

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

Model	Accuracy	Precision	Recall	F1-score
Regresja logistyczna z regularyzacją	0.5862	0.4092	0.5862	0.4625
Regresja logistyczna bez regularyzacji	0.5877	0.4331	0.5877	0.4866
Naiwny klasyfikator bayesowski	0.3051	0.5529	0.3051	0.3737

Sieć neuron.	0.6686	0.7140	0.6686	0.6007
--------------	--------	--------	--------	--------

## Wnioski

Model sieci neuronowej wydaje się być najlepszym z testowanych modeli osiągając najwyższe wyniki we wszystkich metrykach. Regresja logistyczna z i bez regularyzacji uzyskała podobne wyniki, co może oznaczać, że regularyzacja nie jest kluczowa w tym przypadku. Naiwny klasyfikator bayesowski ma niską dokładność ale wyższą precyzję niż modele regresji logistycznej.