

# Raport

06-DUMAUI0 2021/SL

## Cel projektu

Celem projektu było stworzenie modelu, który przewiduje ocenę filmu na podstawie jego roku wydania, gatunku, czasu trwania filmu, ilości głosów, przychodu oraz oceny krytyków (metascore)

## Dane

Dane pochodzą z pliku IMDb movies.csv datasetu „IMDb movies extensive dataset” na platformie Kaggle (link: <https://www.kaggle.com/stefanoleone992/imdb-extensive-dataset> )

Po odrzuceniu obserwacji z pustymi polami pozostało 11316 przykładów, których kolumny znormalizowano do wartości między 0 a 1, rozdzielono kolumnę z gatunkami filmowymi w taki sposób, aby każdy gatunek filmowy miał własną, a następnie podzielono całość na zbiór uczący (9052 przykładów) i zbiór testowy (2264 przykładów).

## Modele

W projekcie porównano działanie 4 modeli:

- Regresja liniowa
- Regresja wielomianowa trzeciego stopnia.
- Regresja wielomianowa trzeciego stopnia z regularyzacją.
- Sieć neuronową (Keras) z dwoma warstwami ukrytymi z kolejno 12 i 8 neuronami. Funkcje aktywacji dla 2, 3 i 4 warstwy to kolejno: ReLU, Sigmoid, Softplus

## Ewaluacja

Do ewaluacji wykorzystano RMSE. Wyniki ewaluacji przedstawia poniższa tabelka:

Model	RMSE train	RMSE test
Regresja liniowa	0.003602	0.003376
Regresja wielomianowa bez regularyzacji	0.002472	0.086475
Regresja wielomianowa bez regularyzacji z funkcją skokową Heaviside'a	0.002472	0.007655
Regresja wielomianowa z regularyzacją	0.003318	0.003350
Sieć neuronowa	0.003625	0.003486

## Wnioski

Najlepsze wyniki na zbiorze testowym można zaobserwować dla regresji wielomianowej z regularyzacją. Podczas gdy na zbiorze treningowym regresja wielomianowa bez regularyzacji uzyskała dobry wynik, można zaobserwować tutaj zjawisko nadmiernego dopasowania przez co uzyskano znacznie gorszy wynik dla zbioru testowego. Bez funkcji skokowej Heaviside'a wyniki dla wielomianów wyższego stopnia przekraczały skalę 0-1 co doprowadziło do znacznego wzrostu RMSE. Przy lepszym dobraniu parametrów sieci neuronowej powinna ona być w stanie uzyskać lepszy wynik niż regresja wielomianowa.