

Uczenie maszynowe — Mikołaj Mazur

06-DUMAU10 2023/SL

Cel projektu

Celem projektu było stworzenie modelu, który przewiduje ilość wypożyczonych rowerów w Chicago na podstawie warunków pogodowych.

Dane

Dane pochodzą z dwóch zbiorów danych.

Jeden zawiera wszystkie wypożyczenia rowerów w roku 2021 udostępnione przez Divvy, system wypożycznia rowerów w aglomeracji chicagowskiej.

[<https://www.kaggle.com/datasets/manjitbaishya001/chicago-bike-sharing-data-2021>] – dostęp 30.01.2024

Drugi zawiera informacje pogodowe z wybranej stacji pogodowej. Ze zbioru było konieczne odrzucenie nieistotnych danych, których daty nie pokrywały się z pierwszym zbiorem danych.

[<https://www.ncei.noaa.gov/access/past-weather/chicago>] stacja CHICAGO MIDWAY AIRPORT 3 SW – dostęp 29.01.2024

Po obrobieniu danych z wypożyczeniami, tak aby jednostką atomową był dzień, złączeniu datasetów oraz sprawdzeniu wartości NaN nastąpił podział na zbiór uczący i testowy w proporcjach odpowiednio 0.75:0.25.

Modele

W projekcie porównano działanie 2 modeli:

- Regresja liniowa. Wraz z normalizacją.
- Regresja wielomianowa trzeciego stopnia. Wraz z normalizacją.

Ewaluacja

Do ewaluacji wykorzystano metryki *Mean Squared Error*, *Mean Absolute Error* i *Root Mean Squared Error*. Wyniki ewaluacji przedstawia poniższa tabela:

Model	Mean Squared Error	Mean Absolute Error	Root Mean Squared Error
Regresja liniowa	16762500	3097	4094
Regresja wielomianowa trzeciego stopnia	85335810	4741	9237

Wnioski

Model z regresją wielomianową, bazując na zbadanych metrykach, okazał się nieco gorszy niż model wykorzystujący samą regresję liniową. Może być to spowodowane pojawiającym się wynikiem odstającym, który w znaczący sposób pogarsza wynik.