

Praca Doktorska

mgr inż. Justyna Dziecioł

Analiza możliwości wykorzystania algorytmów uczenia maszynowego w predykcji współczynnika filtracji dla wybranych kruszyw pochodzenia antropogenicznego

Analysis of the possibility of using Machine Learning algorithms in the prediction of the permeability coefficient for selected aggregates of anthropogenic origin

Praca wykonana pod kierunkiem

Promotora
dr hab. inż. Wojciech Sas, prof. SGGW
Centrum Wodne SGGW

oraz Promotora pomocniczego:
dr inż. Andrzej Głuchowski
Centrum Wodne SGGW

Warszawa, rok 2023



Instytut Inżynierii
Lądowej

Streszczenie

Analiza możliwości wykorzystania algorytmów uczenia maszynowego w predykcji współczynnika filtracji dla wybranych kruszyw pochodzenia antropogenicznego

Aktualna złożona sytuacja branży budowlanej stawia wyzwania związane z odpowiedzialnym gospodarowaniem i zarządzaniem zasobami. Odnosi się to w szczególności do dostępności i pozyskiwania materiałów budowlanych. Pozwoliło to w ostatnich latach na intensyfikację badań mających na celu lepsze poznanie i rozwój zastosowań w budownictwie kruszyw antropogenicznych takich jak żuźle wielkopieczowe, żuźle paleniskowe oraz kruszywa betonowe. W niniejszej pracy zaprezentowano wyniki rozważań nad predykcją współczynnika filtracji z zastosowaniem algorytmów uczenia maszynowego. Współczynnik filtracji jest parametrem istotnym z punktu widzenia zastosowania w geotechnice, inżynierii drogowej czy do oceny stateczności skarp. Wykonano laboratoryjne badania współczynnika filtracji oraz właściwości fizycznych badanych kruszyw naturalnych (piaski) i antropogenicznych (żuźel paleniskowy i wielkopieczowy, oraz kruszywo betonowe). Na podstawie zgromadzonej bazy danych wybrano do analizy algorytmy, które spełniały kryteria zastosowalności do tego rodzaju danych. Dalsze modelowanie przeprowadzono z zastosowaniem algorytmów *AdaBoost*, *Random Forest*, *Gradient Boosting*, *Neural Network*, *k-NN*, jako algorytm kontrolny zastosowano regresję liniową. Zastosowano dwie metody podziału próby: 70% - zestaw uczący, 30% - zestaw testowy, oraz metodę walidacji krzyżowej. Obie z zastosowanych metod podziału próby dały porównywalne wyniki analizy błędów oraz dopasowania R^2 , jednak były one mniej optymistyczne dla metody walidacji krzyżowej. Algorytmem, który uzyskał najlepsze rezultaty dopasowani był *AdaBoost*, którego R^2 wynosił od 0.968 do 0.998. Dla wszystkich algorytmów wykonano szereg analiz statystycznych mających na celu sprawdzenie poprawności szacowanych modeli. Na koniec wykonano analizę interpretacji, jej celem było znalezienie cech - właściwości fizycznych kruszyw mających istotny wpływ na predykcję poszczególnych modeli.

Słowa kluczowe – *współczynnik filtracji, predykcja parametrów geotechnicznych, algorytmy uczenia maszynowego, zrównoważony rozwój, gospodarka w obiegu zamkniętym, regresja liniowa, AdaBoost, Random Forest, Gradient Boosting, Neural Network, k-NN.*

Summary

Analysis of the possibility of using Machine Learning algorithms in the prediction of the permeability coefficient for selected aggregates of anthropogenic origin

The current complicated situation of the construction industry poses challenges in terms of responsible management and stewardship of resources. This is particularly true with regard to the availability and acquisition of construction materials. This has allowed, in recent years, to intensify research aimed at better understanding and development of applications in construction of anthropogenic aggregates such as Blast Furnace Slags, Bottom Slags and Concrete Aggregates. This thesis presents the results of considering the prediction of the coefficient of permeability using Machine Learning algorithms. The filtration coefficient is an important parameter for use in geotechnics, road engineering or for evaluating slope stability. Laboratory tests of the coefficient of permeability and physical properties were carried out for the tested natural (sands) and anthropogenic aggregates (Bottom and Blast Furnace Slag, and Concrete Aggregate). Based on the collected database, algorithms that met the criteria for applicability to this type of data were selected for analysis. Further modeling was performed using *AdaBoost*, *Random Forest*, *Gradient Boosting*, *Neural Network*, *k-NN* algorithms; *linear regression* was used as a control algorithm. Two sample allocation methods were used: 70% - teaching set, 30% - test set, and cross-validation method. Both of the sample partitioning methods used gave comparable results for error analysis and R² fit, but they were less optimistic for the cross-validation method. The algorithm with the best fitting results was *AdaBoost*, whose R² ranged from 0.968 to 0.998. A series of statistical analyses were performed for all algorithms to validate the estimated models. Finally, an interpretation analysis was performed, its objective was to find characteristics - physical properties of aggregates that have a significant impact on the prediction of individual models.

Keywords – Coefficient of permeability, geotechnical parameter prediction, Machine Learning algorithms, sustainability, Circular Economy, Linear Regression, AdaBoost, Random Forest, Gradient Boosting, Neural Network, k-NN.