COMPARISON OF THE POTENTIAL OF APPLYING MACHINE LEARNING METHODS TO PREDICT THE QUALITY OF SURFACE WATER AND TREATED WASTEWATER

M. Piłat-Rożek, E. Łazuka, G. Łagód

Lublin University of Technology, Lublin, PL

Abstract: An electronic nose system consisting of a gas sensor array together with software to analyse multidimensional data from these sensors is used for many environmental engineering applications.
In this work, identical machine learning models were applied to analyse multidimensional data concerning water and wastewater of different quality – surface water originating from drainage system and river as well as different treatment stages in a wastewater treatment plant. The aim of this paper was to compare the application of a PCA method, a Kohonen Self-Organising Map with boundaries superimposed with an unsupervised method, a random forest model and a MLP network on data from different objects. The performance of the mentioned methods, both supervised and unsupervised, was much higher for the data from the wastewater treatment plant. With the algorithms, a clear clustering of the data was observed already with the unsupervised methods alone, while the supervised algorithms classified the observations into the corresponding treatment stages on the training and test sets. The application of the selected unsupervised methods to the drainage and river water data did not allow the separation of distinct groups of observations. Nevertheless, the supervised methods showed greater potential for drawing conclusions – the vast majority of observations were classified into the correct location in the drainage or river. However, the accuracy of the supervised models trained for data from treatment plant is higher than for those from drainages and river.
Keywords: surface water, wastewater, multidimensional data analysis, machine learning