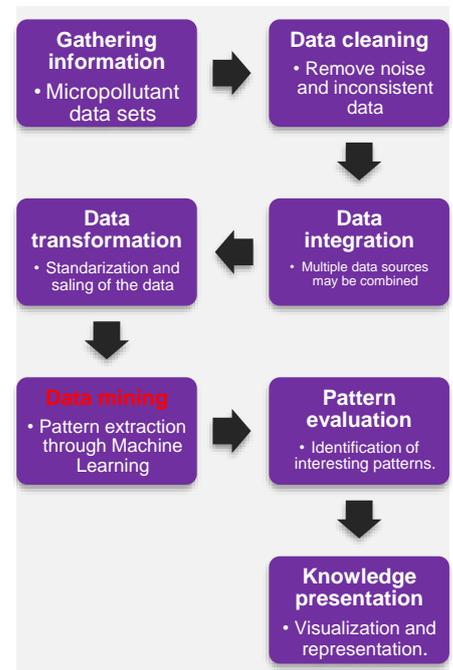


Machine learning techniques applied to the analysis of micropollutants in wastewater: A Data Mining approach.

Background

Contamination of natural water by thousands of chemical compounds despite, for most of them, very low concentrations ($\text{pg}-\mu\text{gL}^{-1}$), raises considerable ecological issues and is a major public concern almost all around the world. Many of these so called ‘micropollutants’ have an urban origin and are used daily in homes, workplaces, or in the urban environment. The biochemical and physical processes applied for the removal of these compounds have been studied previously and further analyzed. However, when handling with huge amount of information (large data sets), non-obvious information within the data is unexploited. The choice of suitable techniques for the data analysis of a large amount of data and multiple variables (>20) demand advanced statistical techniques. Data mining comprises essential processes where intelligent methods (mostly Machine Learning combined with statistical methods based methods) allows us to extract data patterns or unknown information within the data sets. The aim of this study is to provide a complete data analysis through Machine learning techniques of large data sets in order to discover unknown knowledge within the data sets.



Tasks and Requirements

The student will work with large data sets; therefore, at least a basic knowledge of Excel is required. Programming skills are also desirable but not obligatory. The main software for the development of the work will be R (R Foundation for Statistical Computing, 2016). The specific tasks for this study comprise:

- i) A literature background regarding the biochemical (or physical) processes involved in the micropollutant removal.
- ii) A thorough study of Machine Learning techniques applied for discovering unknown patterns within data sets.
- iii) Afterwards, the student should be capable to develop the data analysis for the data sets assigned and together with the theory in the topic, be able to analyze the results found after applying the machine learning methods chosen.

To be considered:

- An introduction to R programming will be imparted by the supervisor and support with the handling of R packages for Machine Learning.

Contact

Dipl. Ing. Luz Alejo
Technische Universität Darmstadt, Institut IWAR
E-Mail: l.alejo@iwar.tu-darmstadt.de
Office: L5|01- 202