

Machine Learning for Predicting and Analysing Water Quality

Daniel Szczypior, Andrew Shields, Dr. Pat Doody

Institute of Technology Tralee, Co Kerry, Ireland

Introduction

One of the most critical issues faced by modern society is quality of natural water resources. In Ireland 81.9% of drinking water comes from surface water (rivers and lakes), 10.5% from groundwater and 7.6% from springs. A key operation for water supply companies is to provide safe drinking water. Therefore, water companies are continually monitoring water quality with various wireless sensors thus generating massive amounts of data, typically stored in data warehouses.

Research Objective

Water treatment systems generally operate to a high level and our water is safe to drink. However, there may be times when significant events occur that impact on water quality. Because of the infrequency of such events they may be classed as anomalies which makes them difficult to detect or predict using standard data mining/monitoring techniques due to the imbalance between the normal and anomaly data.

Methodology

This research proposes the use of data acquired from the Environmental Protection Agency of Ireland (EPA) generated through the Water Framework Directive (WFD) monitoring programme. This project discusses and proposes a solution for data pre-processing and dealing with missing data, it also proposes how to recognise anomalies also called events in the data by the use of EU water regulations to check if the values are within the allowed range.

Potential Outcomes

Using this monitoring data, it is proposed that models capable of predicting water quality may be developed using machine learning techniques. We propose to evaluate the following techniques, Logistic Regression, Linear Discriminant Analysis, Support Vector Machines (SVM), Deep Neural Network (DNN), Time-series Analysis, Artificial Neural Network (ANN), Long short-term memory (LSTM) and Recurrent Neural Network (RNN), which have their performance evaluated by the use of F-score metric.

Data Pre-processing

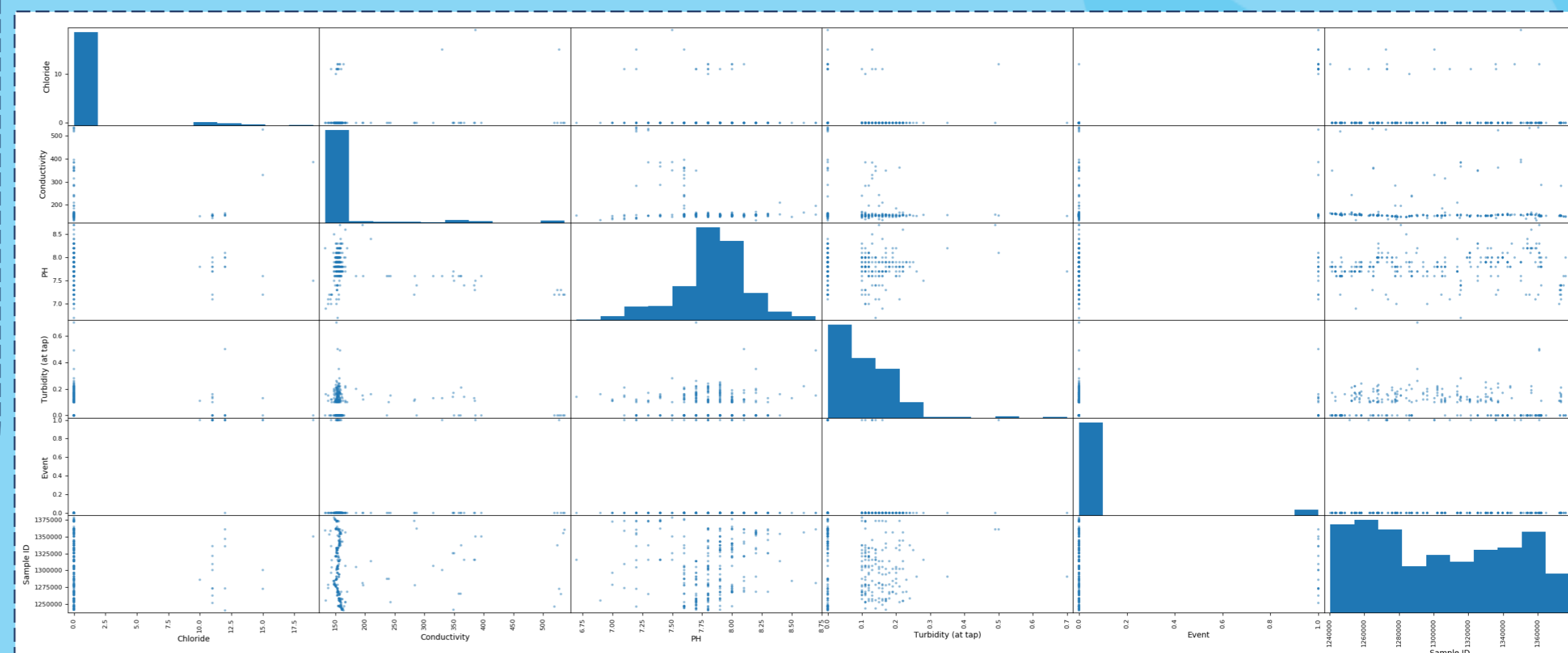
Real world data are generally incomplete, may lack certain relevant attributes, contain only aggregated data or may lack attributes values. They can contain errors, outliers or be inconsistent in naming or code. There are many options we could consider when replacing a missing value, for example: Imputation Using k-NN, Imputation Using Multivariate Imputation by Chained Equation (MICE), Imputation Using Deep Learning (Datawig). The simplest strategy for handling missing data is to remove records that contain a missing value, but it can be too limiting on some predictive modelling.

Event Identification:

Event means that water was polluted, that something exceeded the allowed values. Unfortunately data did not contain any information regarding wherever there was event or not and this is crucial to know for an machine learning algorithm so researching EU and Irish regulations for every of 65 quality parameters was necessary. When it detects that some parameter exceeds allowed range, it sets it as an event.

Preliminary results

Preliminary results are promising, currently finished models are capable of making accurate predictions of water quality. The models were assessed and ordered by the use of their F1 score. SVM scoring second place while Logistic Regression the first. On a scatter matrix below we can see relations between each water parameters and their impact.



REFERENCES

1. Water_Team, E. "Drinking Water Monitoring Results and Water Supply Details for Ireland - Year 2017". Associated datasets and digital information objects connected to this resource are available at: Secure Archive For Environmental Research Data (SAFER) managed by Environmental Protection Agency Ireland <http://erc.epa.ie/safer/resource?id=a0889eb7-6e74-11e8-b825-005056ae0019> (Last Accessed: 2019-02-18)
2. Muharemi, F., Logofătu, D., & Leon, F. (2019). Machine learning approaches for anomaly detection of water quality on a real-world data set. *Journal of Information and Telecommunication*, 3(3), 294–307. <https://doi.org/10.1080/24751839.2019.1565653>
3. S.I. No. 122 of 2014 — (2014). EUROPEAN UNION (DRINKING WATER) REGULATIONS 2014, (122). Retrieved from <http://www.irishstatutebook.ie/pdf/2014/en.si.2014.0122.pdf>

CONTACT

E-mail: Daniel.Szczypior@research.ittralee.ie

Website: <http://www.ittralee.ie>