

Literature Survey on the topic – “Efficient water quality analysis & prediction using Machine learning”.

Survey: 01

Topic: Predicting and analysing water quality using Machine Learning: A comprehensive model.

Authors: Yafra Khan & Chai Soo See.

Published in: 2016 IEEE Long Island Systems, Applications and Technology Conference (LISAT).

Publisher: IEEE

Description: The deteriorating quality of natural water resources like lakes, streams and estuaries, is one of the direst and most worrisome issues faced by humanity. The effects of un-clean water are far-reaching, impacting every aspect of life. Therefore, management of water resources is very crucial in order to optimize the quality of water. The effects of water contamination can be tackled efficiently if data is analysed and water quality is predicted beforehand. This issue has been addressed in many previous researches; however, more work needs to be done in terms of effectiveness, reliability, accuracy as well as usability of the current water quality management methodologies. The goal of this study is to develop a water quality prediction model with the help of water quality factors using Artificial Neural Network (ANN) and time-series analysis. This research uses the water quality historical data of the year of 2014, with 6-minutes time interval. Data is obtained from the United States Geological Survey (USGS) online resource called National Water Information System (NWIS). For this paper, the data includes the measurements of 4 parameters which affect and influence water quality. For the purpose of evaluating the

performance of model, the performance evaluation measures used are Mean-Squared Error (MSE), Root Mean-Squared Error (RMSE) and Regression Analysis.

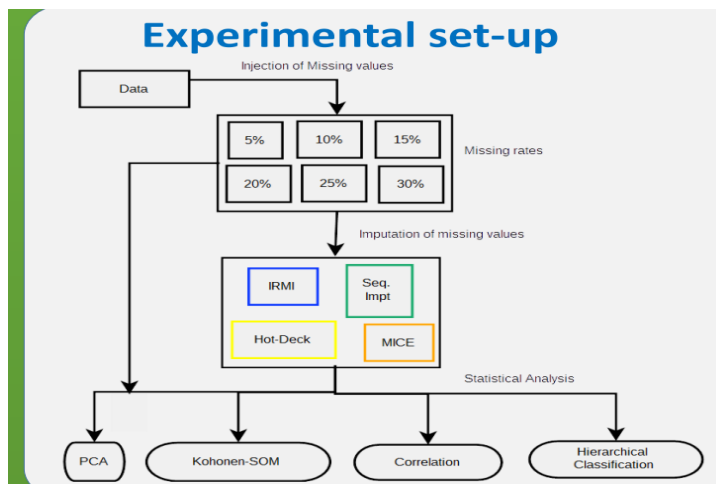
Survey: 02

Topic: Water Quality Data Analytics.

Authors: Eva Serrano, Laure Berti-Equille, Jean-Christophe Desconnets, Maria Aurora Armienta.

Published in: iEMSs. International Environmental Modelling and Software Society Conference.

Description: Water quality monitoring is a regular practice to assess the presence of pollutants in the water. The importance of monitoring is justified by the need to know the current state of aquatic ecosystems to design appropriate conservative and protective actions (Serrano Balderas et al., 2015). Data from water quality monitoring may be prone to have various problems (i.e., incomplete, inconsistent, inaccurate, or outlying data) that may result in misleading analysis interpretation (Berrahou et al., 2015). Incomplete data for instance, can be replaced by imputed values so that the statistical methods commonly used to describe patterns on water quality assessment (such as PCA, Hierarchical Classification, Kohonen-SOM) can be achieved. But imputation of missing values may impact statistical results. In this study, our goal is to assess the impact of imputation methods, and more generally of pre-processing, on the results of various statistical analyses. To this purpose, we studied five imputation methods (Mean, Hot-Deck, Sequential Imputation, Multiple Imputation and Iterative Stepwise Regression Imputation) on four statistical methods (Correlation, PCA, Kohonen-SOM and Hierarchical Classification) and developed a fully integrated analytics environment in R for statistical analysis of environmental data



in general, and for water quality data analytics in particular. The results obtained indicated that the imputation methods IRMI and MI generally improve the accuracy of the tested statistical methods when compared to methods without imputation. Our findings demonstrated that reliable results could be obtained when robust imputation methods are used to pre-process incomplete data.

Survey: 03

Topic: Data-Driven Water Quality Analysis and Prediction: A Survey.

Authors: Jerry Gao.

Published in: The 3rd IEEE International Conference on BIG DATA COMPUTING SERVICE AND APPLICATIONS.

Description: Water quality becomes one of the important quality factors for the quality life in smart cities. Recently, water quality has been degraded due to diverse forms of pollution caused by disposal of human wastes, industrial wastes, automobile wastes. The increasing pollution affects water quality and the quality of people's life. Hence, water quality evaluation, monitoring, and prediction become an important and hot research subject. In the past, many environmental researchers have dedicated their research efforts on this

subject using conventional approaches. Recently, many researchers begin to use the big data analytics approach to studying, evaluating, and predicting water quality due to the advances of big data applications and the availability of environmental sensing networks and sensor data. This paper reviews the published research results relating to water quality evaluation and prediction. Moreover, the paper classifies and compares the applied big data analytics approaches and big data-based prediction models for water quality assessment. Furthermore, the paper also discusses the future research needs and challenges.