

Analysis and Solutions for Water Impurity based on Data Mining

Shreya Bhandge
Department of Computer
Engineering
Sanjivani College of
Engineering, Kopargaon
SPPU, India.

A. A. Muzumdar
Department of Computer
Engineering
Sanjivani College of
Engineering, Kopargaon
SPPU, India.

Charushila Bharambe
Department of Computer
Engineering
Sanjivani College of
Engineering, Kopargaon
SPPU, India.

Nikhil Bachhav
Department of
Computer Engineering
Sanjivani College of
Engineering, Kopargaon
SPPU, India.

Supriya Bakale
Department of Computer Engineering
Sanjivani College of Engineering,
Kopargaon
SPPU, India.

ABSTRACT

It is necessary to digitize the Government for making India digitized especially by making water impurities solution providing digitalized. E-water records i.e. datasets available on the internet that are e-government related services which makes the need of database fulfill and also the data of water impurity more efficient to build or to train the system using machine learning as it is real time records which contains the impurity records that are actually present in India . The system will provide quick solutions with the help of previous training on it with the help of database and will form clusters based on type of impurity type. The training done to system is using neural networks. The solution will be provided by email services.

General Terms

Backpropagation algorithm, k-means +.

Keywords

Neural Networks, Data Mining, Classification, Clustering.

1. INTRODUCTION

In India, it is observed that, there is need for the implementing the solutions for water impurities which will ultimately cause the reduction of human hazards and even avoids affecting on human health. So there is solution that water is tested for various impurities like TDS (Total Dissolved Solids), Turbidity, Nitrate, Fluoride, Chlorine, etc. so these impurities are removed with different solutions if they cross the given threshold which is suitable for drinking purpose.

The village level or taluka level water department concern authorities to their respective district level for testing the water and then applies solution for it. The process is very time consuming process so the propose system is to reduce time consume by giving quick solutions to remove the impurities.

2. LITERATURE SURVEY

The Quality of the water is very important case study in the proposed project. The physical and chemical impurities and test methods of water and limits of impurities as per Indian standards are studied. Natural water or fresh water rarely contains more than 2000 ppm of dissolved solids. Natural

water or fresh water tolerated quantity of alkali carbonates and bicarbonates is less than or equal to 2000ppm. The industrial wastes in water contains inorganic impurities like sulphides, iodides, phosphate, iron, zinc, etc.[1]. The top 10 algorithms for data mining are C4.5, k-means, SVM, Apriori, EM, pagerank, KNN, Naïve Bayes algorithm, CART algorithm are studied and selected k-means algorithm for the clustering[2]. The various contaminants that are studied, its testing is studied in this paper. In the laboratory testing nowadays done is using a scanning fluorimeter. Fluorescence has been used to measure water quality for many years to measure the qualities parameters, from protein, yellow substance to chlorophyll and algae to hydrocarbon and optical brightening agent. The new test case of different coffee brand is used to test and same result is obtained as that of propylene glycol based antifreeze. The use of multiple wavelength devices should be allowing greater differentiation between contaminants [3]. The various impurities that are present state wise and some are with season wise is explained in the paper [4]. The presence of excess of concentration of fluorides, ironand manganese removal is explained with some simple technologies [5].

3. EXISTING SYSTEM

Today in India, the government data of water impurity which is tested for the impurities present in water includes many impurities like fluoride, nitrate, TDS, pH, etc. The data is tested for impurities present in it in district level water department laboratory and then tested datasets and the records of impurity present in water is given to the village level or taluka level government. Then if they have solution for it they apply it in order to remove water impurities present in water. But if they don't have its solutions then again that tested datasets are send to the district level or concern higher authority and then the higher concern authority send the solution to remove the impurity and then the village level or taluka level authority implements or apply that solution to remove impurity. The whole process is very time consuming process which requires 2-3 days.

4. PROPOSED SYSTEM

In our proposed system, the solution providing process which is manual and time consuming in existing system is done quickly and by email in proposed. The dataset that is already

present on internet on government website [6]. The datasets contains various impurities that are generally present in water that datasets and its solutions will be feeded to system using neural network. Neural network is the very popular them in data mining which is used for classification and even in regression and in machine training. The system will be then trained with datasets. The Backpropagation algorithm is used. When concern authority gives datasets that is tested and feeded to system to check for its results. The system then will give result whether the water whose records are given as datasets to system is supposed to be drinkable or not. If the water records datasets results that the water is not drinkable then the system will result solution to remove those water impurities. The clusters of impurities are formed according to the type of impurities. That clustering is done using K-means algorithm studied in paper [2]. That solution and clusters is then sent to the water department concern authority. The concern authorities then implements or apply that solution to remove water impurities. The Civilians are also allowed to send their feedback regarding the water they use like its colour, odour, taste, etc. The solution provided by them is then stored in database2 and can be used for further testing. The water department concern authority is provided a solution in less time compared to existing system.

5. SYSTEM ARCHITECTURE



Figure 1: System Architecture

This section contains the information and details of modules present in the above system architecture Figure 1.

Modules: As shown in Figure 1, describes the modules of our proposed system.

Admin Module: This module contains the admin task i.e. the task performed by the admin such as creating the database and deleting the database and also updating the database. Here the security algorithm is provided.

System Training: This module contains the historical data i.e. the system will be trained based on historical data.

Analysis: This module contains the impurity records and also selects the frequent and important parameters and will have cluster for each impurity. Here we will use clustering

algorithm. The analysis of impurities is also done in this module and for that classification we will use Naïve's Bayesian classification.

Solution Finding: This module contains the previous records of water impurities tested or historical data and based on that the solution will be given and if the previous record does not match the new tested impurity record then that record will be added in the database. Here we will use decision tree algorithm named as C4.5 algorithm.

User (employee): It contain the user or the concern authority who will get the database of old and new records of the solution given by the system.

Feedback: This module contain the website which will be for normal user who will give the feedback about the water quality at their home..

6. CONCLUSIONS

We are going to develop the system which will take the huge database as records, and then the operations will be done on the certain selected attributes, to get the results i.e. to find the solution for the given impurity. The solution is given to the concern authority. The perfect solution based on the previous records is given to the user i.e. concern authority. The records as already given and trained to the system and then the analysis and solution is done so the mode is supervised mode in which all the previous data records are already present and hence the operation is done on that, so that to produce desired output i.e. to produce the solution based on the impurity.

7. REFERENCES

- [1] Mr. K. J. Kucche, Dr. S. S. Jamkar, Dr. P.A. Sadgir, "Quality of Water for Making Concrete: A Review of Literature", International Journal of Scientific and Research Publications, Volume 5, Issue 1, January 2015.
- [2] Xindong Wu, Vipin Kumar, J. Ross Quinlan, Joydeep Ghosh, Qiang Yang, Hiroshi Motoda, Geoffrey J. McLachlan, Angus Ng, Bing Liu, Philip S. Yu, Zhi-Hua Zhou, Michael Steinbach, David J. Hand, Dan Steinberg, "Top 10 algorithms in data mining", Springer-Verlag London, 4 December 2007.
- [3] Julius Okach, Barry Haggett, Robin Maytum, Andrew Mead, David Rawson, Tahmina Ajmal, "Sensing fresh water contamination using fluorescence methods", Department of Computer Science and Technology University of Bedfordshire, Luton, United Kingdom.
- [4] M. Dinesh Kumar and Tushaar Shah, "Groundwater Pollution and Contamination in India: The Emerging Challenge".
- [5] Vingeswaram, C. Visvanatham, M. Sundarvadivel, "Treatment Options for removal of specific impurities from water", Waste water recycle reuse, and reclamation-vol2.
- [6] https://www.google.co.in/search?q=www.indiawater.gov.in&ie=utf-8&oe=utf-8&client=firefox-b&gfe_rd=cr&ei=4ADwV4ubFJT8geF-IGYBw