Poll Res. 41 (4) : 1178-1183 (2022)

ISSN 0257–8050

DOI No.: http://doi.org/10.53550/PR.2022.v41i04.005

AN INTELLIGENT DECISION SUPPORT SYSTEM FOR WASTEWATER TREATMENT PLANTS IN THE SULTANATE OF OMAN

SHERIMON PULIPRATHU CHERIAN[1*], ALAA ISMAEEL[1], MANNA ANN MARIYAM[2], SIAN GIJO[2], WINNY ANNA VARKEY[2], SHERIN JOHN[2] AND NISHA JOSEPH[2]

[1]Faculty of Computer Studies, Arab Open University, Muscat, Sultanate of Oman
[2]Department of Computer Science & Engineering, Saintgits College of Engineering, Kottukulam Hills, Pathamuttam P.O. 686532, Kerala, India

(Received 21 June, 2022; Accepted 8 August, 2022)

ABSTRACT

All living things need access to clean water to survive. However, drinkable water is scarce, and as a result of human activity, these supplies are becoming severely contaminated. In order to replenish these depleting water supplies while also reducing contamination-causing activities, several steps must be made. Wastewater treatment plants (WWTPs) are essential for removing toxins from various sectors so that clean water may be released into the environment with the least amount of environmental harm. It involves a combination of complex processes used to treat and remove pollutants from water. All the decisions in WWTPs are conventionally taken by skilled and qualified plant operators with the necessary training and education in order to get the job done right. There can be a considerable amount of error that can occur when critical decisions are taken by these operators. In order to tackle this and to improve efficiency and accuracy, a Decision Support System (DSS) can be used as traditional methods of decision making by human operators are considerably less efficient. Water quality parameters such as pH, hardness, solids, chloramines, sulphate, conductivity, organic carbon, trihalomethanes, turbidity are used to determine the purity status of water being considered. The proposed study focuses on a Machine Learning based DSS built on Decision Tree (DT) algorithm that will predict the purity of water using historical data, which will aid the plant operators in making daily operational decisions at the WTTPs. The experimental result analysis shows that the model built using DT algorithm gives good performance.

KEY WORDS : Intelligent Decision Support System, Machine Learning, Decision Tree

INTRODUCTION

The second-largest nation in the Arabian Peninsula is the Sultanate of Oman. The shoreline extends to 1700 km dotted with several islands offshore. Its ecosystem is unique and is vulnerable to disturbance from various activities (Choudri *et al.*, 2015). Oman's water sector is experiencing strong demands primarily due to the population growth and tourism projects, and they are projected to rise in the near future. The World Bank Collection of development indicators reports that the percentage of contamination in Oman's water is 75.38 % in 2007. The survival of mammal, bird, fish, and plant species in Oman is at risk because of this. The majority of cities in the Sultanate employ holding tanks and septic systems to collect sewage water from residential areas (Jaffar Abdul Khaliq *et al.* 2017). According to recent studies, the dissolved oxygen concentration in the Gulf of Oman's coastal waters is essential and has been steadily declining, posing a major threat to marine life. All communities must make sure they have efficient water treatment systems in place so that treated water may be recycled or utilised again, especially those in water-scarce areas.

In order to restore water supplies and safeguard the ecosystem from contaminants, wastewater treatment is crucial. Both domestic and commercial toxins may be found in wastewater. If wastewater is not treated, the pathogens and chemical compounds it contains can harm birds, plants, and animals that live in or near water. Additionally, it can contaminate food supplies and water supplies, endangering people's quality of life (Farkas *et al.*, 2020). The health of numerous ecosystems depends on the treatment of wastewater (Akpor and Muchie 2011). Wastewater can serve as a source of water for many purposes when it has been treated appropriately. Conventionally, technicians make decisions in Wastewater Treatment Plants (WWTPs). This necessitates advanced technical abilities as well as a thorough understanding of the procedures and steps involved. A minor error in assessment might have far-reaching effects. These decisions are vulnerable to human errors. The importance of wastewater treatment is increasing day by day. If wrong decisions are taken, this can adversely affect the environment into which it is released. Instead, if a single intelligent system is put in place, it can make precise decisions, lowering the potential expense associated with having human errors corrected. The nation's economic health can benefit from this low-cost approach. As a result, an intelligent decision support system is required. With the use of Artificial Intelligence (AI), the suggested system will assist technicians in WWTPs in numerous everyday decision-making processes.

The main objectives of this study are:

- Build an Intelligent Decision Support System (IDSS) for WWTPs using AI by employing a Machine Learning model.
- Create a User Interface for the interaction of the plant operator with the system to generate the desired application.

MATERIALS AND METHODS

AI based Decision Support System (DSS) for WWTPs describes a model that learns from historical data gathered from the treatment plants and uses this knowledge to identify new and unseen data. By predicting the suitability of water based on the relevant water quality metrics that are provided as input to the system, this will assist plant operators in decision-making at the treatment plants. The proposed system utilises this functionality by modelling an AI model built on Decision Tree (DT) algorithm (Song and Ying, 2015). The schematic representation of prediction model is shown in Fig.1.The model takes into account nine attributes- pH, hardness, solids, chloramines, sulphate, conductivity, organic carbon, trihalomethanes, and turbidity- which are water quality criteria that establish the state of the water's purity level. Also, a simple User Interface is created to aid plant operators to enter daily water quality parameter readings.

Water quality attributes

In this work, we considered nine water quality attributesthat control the water quality determination and they are: pH, hardness, solids, chloramines, sulphate, conductivity, organic carbon, trihalomethanes, and turbidity (Desai and Sungkur 2022; Pal, 2022).

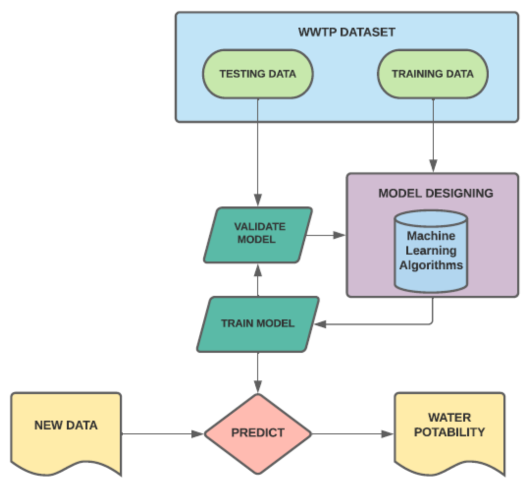

Fig. 1. Schematic representation of prediction model

pH – An indicator of a solution's acidity or alkalinity, pH is a logarithmic scale. Low numbers are more acidic and high ones are more alkaline where 7 is neutral.

Hardness – High mineral content water is referred to as hard water. When water seeps through limestone, gypsum, or other formations predominantly made of calcium carbonate, magnesium, bicarbonate, and sulphate, hard water is the result. Health advantages of hard water are modest.

Solids – It measures the total dissolved content of all inorganic and organic compounds found in a liquid in the form of molecules, ions, or tiny particles that are suspended in the liquid.

Chloramines – Chloramines are disinfectants used to treat drinking water. Chloramines are most commonly formed when ammonia is added to chlorine to treat drinking water.

Sulphate – Sulphates are naturally occurring anions in both surface and groundwater freshwater. Sulphate levels above 250 mg / l can add bitterness to water.

Conductivity – Mineral salts of elements like sodium, calcium, and magnesium are the cause of it. These salts create free ions when they are dissolved in water, which allow the water to conduct electricity.

Organic Carbon– The quantity of organic compounds in a sample of water is determined by the total organic carbon. More oxygen is used when a substance has a higher carbon or organic content. High levels of organic materials promote the growth of microbes that deplete the oxygen supply.

Trihalomethanes – When chlorine compounds used to disinfect water mix with other naturally occurring chemicals in water, a class of disinfection by-products known as total trihalomethanes is created.

Turbidity – Turbidity is a measurement of a liquid's relative transparency. This is a measurement of the quantity of light scattered by elements in water when it is illuminated through a water sample. It is an optical property of water.

Decision treealgorithm

A business or firm's decisions, judgments, and courses of action are aided by an IDSS. Large amounts of data are sorted through and analysed by a data science system, producing complete data that can be used to solve issues and make choices. By expediting process simulation, an essential stage in both building new plants and improving the design of existing ones, machine-learning technologies can aid in wastewater treatment process design. In the proposed method, we use Decision Tree (DT) to build the AI model (Song and Ying, 2015). DT algorithm is a kind of Supervised Learning where the learning takes the form of a tree-structure consisting of decision nodes and leaves as shown in Fig. 2. DT can be used to address classification and regression problems. The name itself suggests that it uses a flowchart that mimics a tree structure to represent the predictions that result from a series of feature-based splits. A decision root node is the node from which the population begins to be divided based on several features. Decision nodes are the nodes that result from separating the root nodes. Leaf nodes or terminal nodes are the nodes where further splitting is not possible. Pruning is little more than removing a few nodes to prevent overfitting. The procedure in a DT starts at the root node and moves upward to predict the class of the input dataset. Based on a comparison of the values of the root property and the record attribute, this method tracks the branch and goes to the next node. The process is then carried out by once more comparing the attribute value of the next node with those of the other sub-nodes. It keeps doing this until it reaches

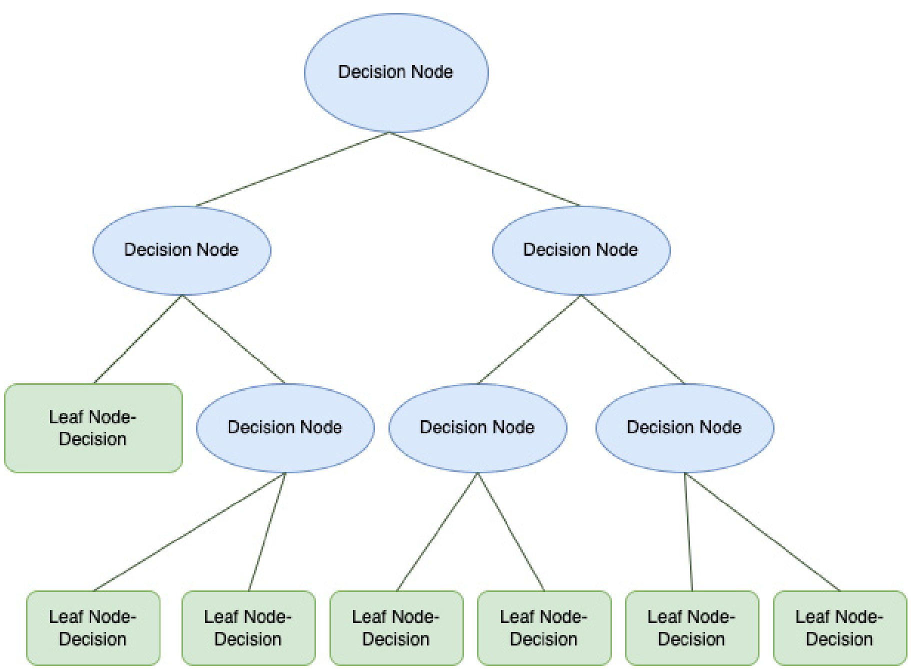

Fig. 2. Decision tree structure

the leaf node of the tree.

When utilising DT to forecast a class label for a record, we begin at the tree's base. The root attribute and the record's attribute values are contrasted. On the basis of the comparison, we move to the next node by following the branch that corresponds to that value. Using DT, the instances are categorised by being arranged from the root of the tree to a leaf or terminal node, with the leaf or terminal node providing the classification. Every node in the tree represents a test case for a particular feature, and every edge descending from the node provides a potential solution for the test case. Each subtree anchored at the new node goes through this cyclical process. The results can be continuous-valued or categorical (for instance, Yes or No).

RESULTS AND DISCUSSION

The details about the experimental setup, data collection, performance evaluation metrics, and experiments and result analysis areexplained in the following subsections.

Experimental setup

The prediction experiments are carried out in Jupyter Notebook (Jupyter Notebook, 2022). A Personal Computer (PC) with an Intel Celeron N3050 processor and 4 GB of RAM has been used to run the simulation and handle all required jobs.The model construction is facilitated by using development environments that offer machine learning standard libraries such as scikitlearn (Scikit-learn, 2022). Scikit learn is interoperable with other libraries such as NumPy and Pandas that together facilitates the development and visualisation of this model.

Data collection

Data for this study is collected from multiple Oman-wide WWTPs. It contains 3,276 samples collected from various locations in the Sultanate of Oman. Nine relevant metrics are present in the dataset: pH, hardness, solids, chloramines, sulphate, conductivity, organic carbon, trihalomethanes, and turbidity. The dataset considered for this study is cleaned by replacing the missing values by the mean of remaining values of that attribute.

Performance evaluation metric

Accuracy is used as the metric to assess the performance of the classification model and it is calculated from the Confusion matrix (Townsend 1971). The confusion matrix is a well-known method for dealing with classification issues. When describing how well a classification model performed on a set of test data for which the true values are known, a confusion matrix is a table that is widely employed. To display counts based on actual and expected values, confusion matrices are utilised. The number of correctly identified negative cases is shown by the output "TN," which stands for True Negative. Similar to this, the abbreviation "TP" stands for True Positive and represents the quantity of correctly detected positive cases. False Positive and False Negative values, respectively, are denoted by the letters "FP" and "FN." False Positive Value (FP) is the number of real negative examples that are mistakenly categorized as positive, while False Negative Value (FN) is the number of real positive examples that are mistakenly classified as negative (Brown, 2018).

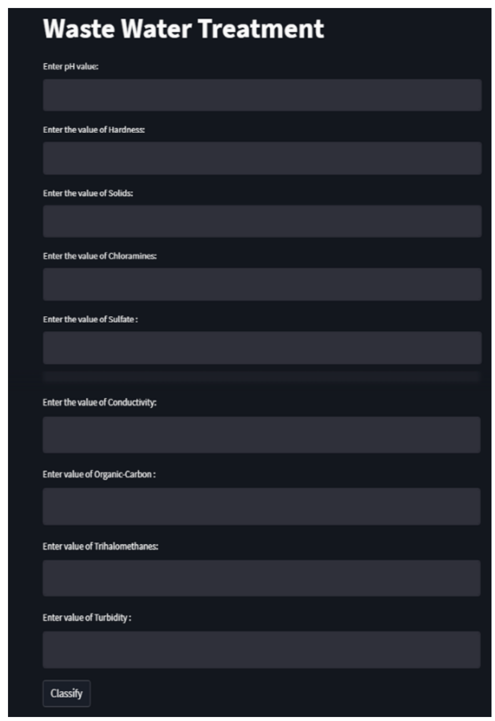

Fig. 3. User interface

Experimental results and analysis

The dataset used is initially pre-processed to handle the null values in the dataset. This is done by replacing the null values by the mean of all the remaining values of that attribute. This pre-processing is done for all the features. The dataset is then divided into testing and validation sets, with

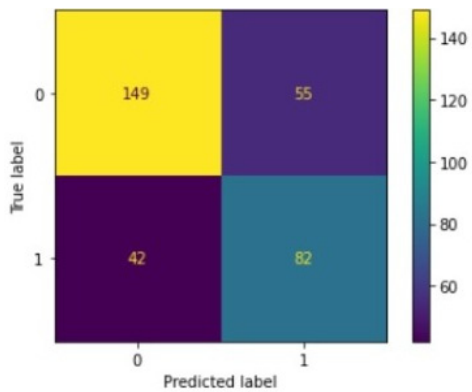

Fig. 4. Confusion matrix of the model

90% going to training and 10% to validation. Using Jupyter notebook, the AI model using DT algorithm is created by loading libraries like Num Py, pandas, scikit-learn, and others.The system also provides a simple and user-friendly interface as shown in Fig.3, that can be used by plant operators with minimal technical knowledge. The user interface is created using Streamlit by Python language which is an open-source web application framework for creating apps for machine learning applications (Streamlit-A faster way to build and share data apps, 2022). There are nine input text boxes where the user can enter water quality parameter readings such as pH, solids, turbidity, etc. and a submit button 'Classify' to submit the input test data to the model. The model then predicts the potability by displaying a binary 0/1 output where a binary '1' indicates that the water being tested is potable and a '0' indicating otherwise.

The confusion matrix of the classification model built by DT algorithm is obtained and as shown in Fig.4. The classification accuracy of the model is calculated using the confusion matrix.

Also, the performance of the model built using DT algorithm is compared with other Machine Learning (ML) algorithms like k-Nearest Neighbour (k-NN)(Kataria and Singh, 2013), Naive Bayes (Rish 2001), Random Forest (Biau and Scornet, 2016) and Support Vector Machine (SVM) (Noble, 2006). The performance comparison is provided in Table 1. Among all the algorithms compared, it is DT algorithm gives the maximum accuracy of about 70.42%. It can be seen from the DT visualisation that the model has chosen the feature Sulphate as the best splitting algorithm and hence forms the root decision node. Subsequent tree growth can be studied from Fig. 5.

CONCLUSION

Wastewater contains hazardous elements; it is required to be treated before it is released into the environment or reused for other purposes. Hence, wastewater treatment proves to be critical. Even though conventional methods proved to be useful in small scale facilities, as the processes become more complex and exhibit nonlinear nature, these methods proved to be insufficient. Earlier decisions in WWTPs solely depended on the knowledge and expertise of the experts and technicians in charge,

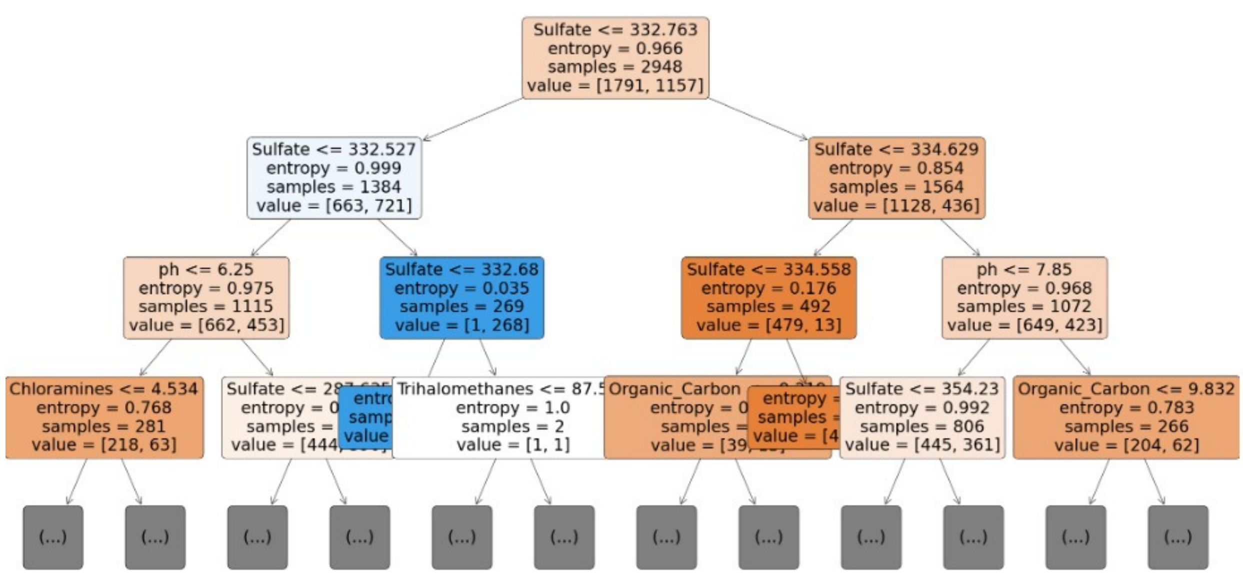

Fig. 5. Decision tree visualisation

and hence itwas error prone. Hence, in this work, we designed an Intelligent Decision Support System using DT algorithm. The proposed model predicts the purity of water using historical data, which will aid the plant operators in making daily operational decisions at the WTTPs. The experimental result analysis shows that the model built using DT algorithm gives good performance.

Table 1. Comparison of accuracy of ML algorithms

Sl No.	Models	Accuracy (%)
1	SVM	59. 45
2	k-NN	64.93
3	DT	**70.42**
4	Random Forest	61.58
5	Naive Bayes	57.16

ACKNOWLEDGEMENT

The research leading to these results has received funding from the Research Council (TRC) of the Sultanate of Oman under the Block Funding Program BFP/RGP/ICT/20/453.

Conflict of Interest

There are no conflicts of interest regarding the publication of this work.

REFERENCES

Akpor, O. B. and Muchie, M. 2011. Environmental and public health implications of wastewater quality. *African Journal of Biotechnology*. 10 : 2379-2387. https://doi.org/10.5897/AJB10.1797.

Biau, Gérard and Erwan Scornet, 2016. A random forest guided tour. *Test*. 25 : 197-227. https://doi.org/10.1007/s11749-016-0481-7.

Brown, J. B. 2018. Classifiers and their metrics quantified. *Molecular Informatics*. 37 : 1-11.

Choudri, B. S., Mahad Baawain, and Mushtaque Ahmed, 2015. Review of Water Quality and Pollution in the Coastal Areas of Oman. *Pollution Research*. 34: 229-239.

Desai, Bhoushika, and Roopesh Kevin Sungkur, 2022. Software Quality Prediction Using Machine Learning. *Lecture Notes in Networks and Systems* 393: 401-411. https://doi.org/10.1007/978-3-030-94191-8_32.

Farkas, Kata, Luke S. Hillary, Shelagh K. Malham, James E. McDonald, and David L. Jones, 2020. Wastewater and public health: the potential of wastewater surveillance for monitoring COVID-19. *Current Opinion in Environmental Science and Health*. 17. Elsevier Ltd: 14-20. https://doi.org/10.1016/j.coesh.2020.06.001.

Jaffar Abdul Khaliq, Suaad, Mushtaque Ahmed, Malik Al-Wardy, Ahmed Al-Busaidi, and Choudri, B. S. 2017. Wastewater and sludge management and research in Oman: An overview. *Journal of the Air and Waste Management Association*. 67. Taylor & Francis: 267-278. https://doi.org/10.1080/10962247.2016.1243595.

Jupyter Notebook, 2022. https://jupyter.org/. Accessed July 21.

Kataria, Aman, and Singh, M.D. 2013. A Review of Data Classification Using K-Nearest Neighbour Algorithm. *International Journal of Emerging Technology and Advanced Engineering*. 3 : 354-360.

Noble, William, S. 2006. What is a support vector machine? *Nature Biotechnology*. 24 : 1565-1567. https://doi.org/10.1038/nbt1206-1565.

Pal, Osim Kumar. 2022. The Quality of Drinkable Water using Machine Learning Techniques. *International Journal of Advanced Engineering Research and Science*. 9 : 016-023. https://doi.org/10.22161/ijaers.96.2.

Rish, Irina, 2001. An empirical study of the naive Bayes classifier. In: *IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence*. 3 : 41-46. https://doi.org/10.1039/b104835j.

Scikit-learn, 2022. https://scikit-learn.org/stable/. Accessed July 21.

Song, Yan-Yan, and Ying, L. U. 2015. Decision tree methods applications for classification and prediction. *Shanghai Archives of Psychiatry*. 27: 130-135.

Streamlit-A faster way to build and share data apps. 2022. https://streamlit.io/. Accessed July 21.

Townsend, J. T. 1971. Theoretical analysis of an alphabetic confusion matrix. *Perception & Psychophysics*. 9 : 40-50. https://doi.org/10.3758/BF03212817.