Scholar Publishing Group
Water Pollution Prevention and Control Project
https://doi.org/10.38007/WPPCP.2022.030401
ISSN 2790-2730 Vol. 3, Issue 4: 1-10

Water Pollution Prevention Engineering Based on K-Means Algorithm and Naive Bayesian Algorithm

Fang Li*

Department of Information Engineering, Heilongjiang International University, Harbin 150025, China

lifang@hiu.net.cn

**corresponding author*

Keywords: Water Pollution, Prevention and Control Engineering, Classification Algorithm, Clustering Algorithm

Abstract: Water resources are important resources for people's life and development. However, sewage not only has a huge impact on human health and environment, but also causes a series of negative impacts. In the process of social development, paying attention to the prevention and control of sewage and adopting various effective prevention and control measures can effectively protect water resources and create favorable conditions for the coordinated development of human and nature. Many researchers have provided new ideas for the application of Water Pollution (WP) prevention and control system, and this is the research direction and basis of this paper. This paper analyzed the significance of WP prevention and summarized the research on WP prevention and control engineering and related naive Bayesian algorithm. Then the algorithm model was established, and the relevant algorithms were proposed to study the WP prevention and control engineering based on K-Means algorithm and naive Bayesian algorithm. At the end of the article, the simulation experiment was carried out, and the experiment was summarized and discussed. Firstly, the selected areas were analyzed. The number of WP prevention projects in cities would increase from 2018 to 2021. However, the number of WP prevention and control projects in rural areas has increased from 2018 to 2020, which meant that WP prevention and control work in rural areas was generally easier to carry out than in cities. Through the comparison of natural and social conditions in Region A and B, it was concluded that the difficulty of prevention and control projects was also determined by areas with poor natural original conditions but exceptionally developed social conditions. At the same time, with the in-depth study of K-Means algorithm and naive Bayesian algorithm, the application research of WP prevention engineering is also facing new opportunities and challenges.

1. Introduction

With the advancement of urbanization, people are facing the increasing pressure of water resources shortage. The deep exploration of the root of urban water crisis has gradually made people realize that the social cycle of water has exceeded the limit that the natural cycle of water can bear. It is an effective way to realize the sustainable development of water resources to reasonably use water resources on the premise of fully understanding the natural movement law of water, and ensure that the upstream water cycle does not affect the function of the downstream and does not affect the social cycle. This requires a shift from a single and open approach to a feedback cycle in order to make better use of water resources.

Research on WP prevention and control engineering. Wang Yubao has studied the establishment of environmental complaint reporting systems in many regions, and the real-time monitoring and management of industrial WP [1]. The application and sustainable development of biofuels in urban wastewater treatment, industrial wastewater treatment and rainwater treatment were discussed by He Mingjing [2]. The results of Long Yan's research showed that the algae and crustaceans in the middle part of the South-to-North Water Diversion Project are in a very safe state, effectively solving the shortcomings of the existing assessment methods [3]. Hou Wei's research results provided comprehensive information to better understand the local and global pollution levels, and may help to prevent and control pollution [4]. Lakshmikantha Varsha believed that with the increasing pollution of drinking water, WP is one of the main threats in recent years. Polluted water would cause various diseases to humans and animals, thus affecting the life cycle of the ecosystem [5]. Liu Xian believed that the complexity and dynamics of the environment make it extremely difficult to directly predict and track the time and space changes of pollution [6]. The above studies have achieved good results, but there are still some problems with the continuous updating of technology.

The naive Bayesian algorithm is studied in WP prevention and control engineering. Meng Qingxuan believed that it is important to identify and correct incomplete water quality data. A data cleaning method based on improved balanced iterative reduction and hierarchical clustering algorithm was proposed [7]. Kumar K Kishore's research on plant image detection based on machine learning has been widely developed in the agricultural field [8]. Lodhi Pooja believed that pollution increased with the increase of human water use. In order to control the effects of the pollutants and take measures to reduce WP, he proposed several methods. Water quality index measurement is a method used to measure the composition of harmful water [9]. The results of the Rastogi Rohit's assessment confirmed the ability of the proposed framework to make continuous geographical observations of air quality. In addition, air quality in the entire population can be continuously monitored [10]. Wu Jianhua have used correlation analysis, principal component analysis, hierarchical clustering analysis, and multidimensional linear regression to investigate the sources and factors of WP subsurface effects [11]. Nguyen, Hiep Duc studied that the modeling system has proved to be a valuable tool for predicting the impact of different water sources in urban areas on water quality in waterway areas that are critical to the protection of ecosystems and human health [12]. The above research shows that the application of K-Means algorithm and naive Bayesian algorithm has a positive effect, but there are still some problems.

This paper studies the research of WP prevention and control engineering based on K-Means algorithm and naive Bayesian algorithm. Firstly, the WP prevention and control project is explored, and then its relevant content is given. The application of clustering and classification algorithm in WP prevention and control project is designed, and the relevant algorithm provides theoretical basis for the experiment. Finally, the WP prevention and control systems in the selected areas are compared and analyzed under the K-Means algorithm and naive Bayesian algorithm, providing

reference significance for such research.

2. Evaluation on WP Prevention and Control Project

2.1. Evaluation of WP Prevention

WP prevention and control is an environmental engineering technology closely related to local natural and social conditions. Therefore, it is necessary to carry out overall planning and comprehensive regulation from various aspects. According to the planning of the river basin, the environmental capacity and self-purification law are studied in sections and zones, and the pollution load is determined. The pollution prevention plan is formulated and the pollution source is treated, including the establishment of regional joint sewage treatment plant, water regulation, pollution load, etc. The self-purification performance and water quality can be improved by diverting water from adjacent water systems and setting aeration facilities. In addition, it is necessary to strengthen the treatment of surface runoff such as farmland and mines, improve various wastewater treatment technologies and treatment efficiency, and save energy. It is also necessary to rationally use water resources, and realize wastewater recycling and other measures to achieve comprehensive treatment of WP [13].

2.2. Prevention and Control of WP Hazards in the Project

The following summarizes the hazards of WP in relevant prevention and control projects, as shown in Figure 1:

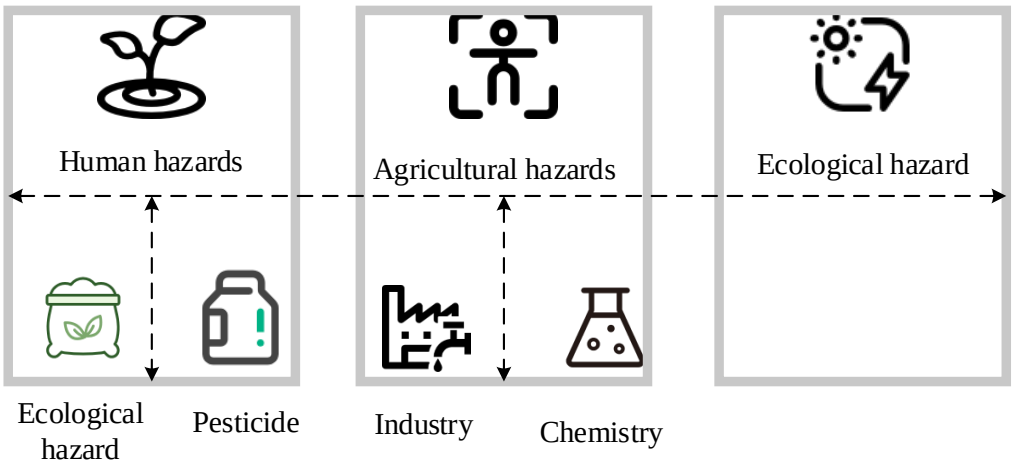

Figure 1. Hazards of water pollution in the prevention and control projects

2.2.1. Human Hazard

Water is the source of life. Unsanitary water would cause harm to human body. According to scientific research, many diseases are caused by substandard water quality. If harmful substances are accumulated in the body for a long time, they would cause serious harm to the body. There are many toxic pollutants in industrial wastewater, and people would have toxic reactions after drinking. With the rapid development of industry and economy, the problem of WP is becoming increasingly serious and gradually changing to chemical pollution. Due to the lack of attention paid by the public to health care, some areas may even suffer from collective poisoning caused by WP. At present, most organic compounds are highly toxic and pollute the water source, but it is difficult to achieve

the purification effect. Long-term drinking of contaminated water would lead to the continuous accumulation of toxins in the body and eventually lead to a variety of diseases [14].

2.2.2. Agricultural Hazards

If the underground water source is polluted and used for production and agriculture, it would lead to corrosion and damage of equipment, which would have a certain impact on agricultural production and reduce the service life of equipment. In addition, if chemical fertilizers, pesticides and other chemicals are used in the agricultural production process, the water from farmland irrigation would be irrigated into the soil, thus polluting the groundwater and leading to the decline of agricultural production and quality. At the same time, in the process of irrigation, if the contaminated water is used for irrigation, it would not only lead to changes in soil structure, but also affect the growth of crops, affect the resistance of crops and thus affect agricultural production.

2.2.3. Ecological Hazard

As far as the current situation is concerned, the most serious situation is irreparable damage to the ecological environment of the water body. A large amount of domestic sewage and industrial wastewater would affect the purification function of the water body, and would also produce some organic substances. These organic substances would cause a large number of algae in the water body to multiply, resulting in a decrease in the oxygen content in the water body. Oxygen is an important substance for aquatic animals to survive and also an important substance for purifying pollution. If the oxygen concentration is insufficient, a large number of marine organisms would die, causing serious damage to the ecological environment [15].

2.3. Control Measures for WP in the Project

The following summarizes several treatment measures for WP in the project, as shown in Figure 2:

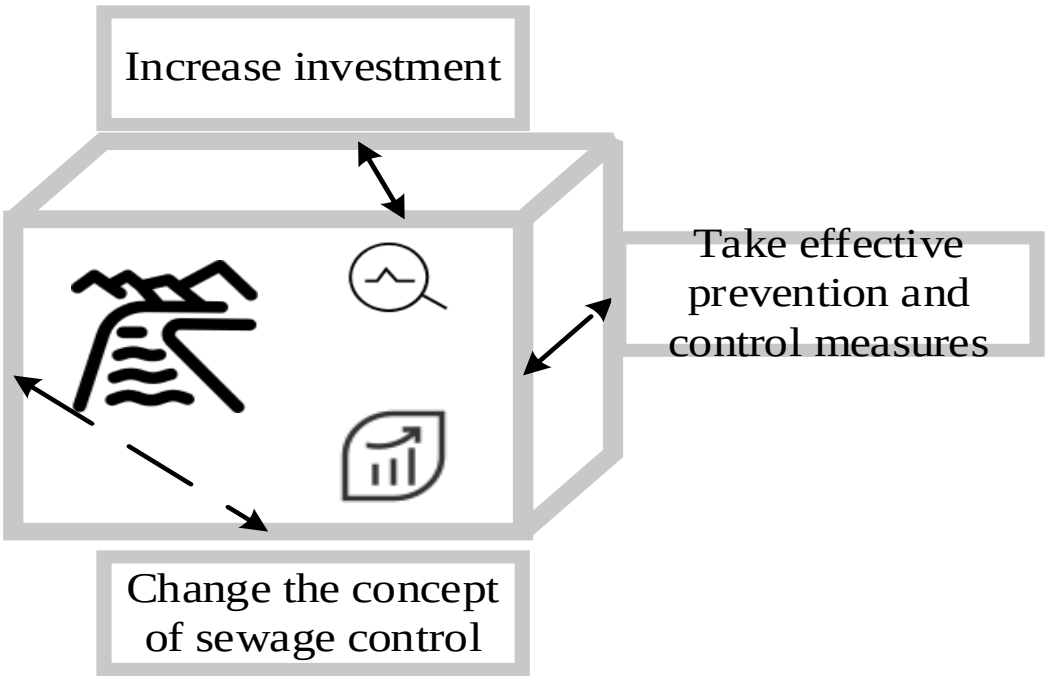

Figure 2. Control measures for water pollution prevention and control in projects

2.3.1. Change the Concept of Sewage Control

In order to improve the level of urban sewage treatment, people should change ideas and strengthen the treatment of urban domestic sewage. Governments at all levels should formulate relevant sewage treatment laws and regulations according to the WP situation in different regions to ensure that they are compatible with the local reality, and incorporate them into urban sewage treatment, wastewater recycling and strict control. It is necessary to connect cities and sewers, improve the working environment of relevant departments and government departments, and strengthen the treatment of urban sewage. At the same time, people should improve the current public relations project, increase the publicity of the water environment, regularly set up sewage treatment towers in the government's publicity, strengthen the guidance of urban ecological publicity and deepen the understanding of the theory through strengthened training.

2.3.2. Taking Effective Prevention and Control Measures

As the sewage treatment process and flow are relatively complicated, targeted treatment methods should be adopted according to the actual situation of different regions. At the same time, it is necessary to strengthen the supervision and allocate corresponding staff, carry out regular or irregular supervision on each link, timely grasp the implementation of other links, and analyze the implementation of other links and put forward corresponding countermeasures. In addition, in order to improve the effect of sewage treatment and the quality of sewage treatment, people must introduce social control and implement sewage treatment in sewage treatment. Relevant departments should actively use the Internet and other media to publicize the importance of sewage treatment, and make the connection point between sewage treatment and target achievement public on the Internet and strengthen the monitoring efficiency.

2.3.3. Increasing Investment

If people want to improve the efficiency of sewage treatment, people must increase investment in capital, technology and manpower to ensure the normal development of sewage treatment. Due to the relatively small investment of relevant departments in the treatment of urban domestic sewage, in order to better treat urban wastewater, people must constantly carry out reform and keep open in the financial environment. The managers of pollutant discharge enterprises should have new ideas, innovate and develop the business model of credit funds, and gradually expand the scope of urban pollution discharge and use. While raising funds for sewage treatment, relevant departments should also strengthen investment in sewage treatment.

3. Application of Clustering and Classification Algorithm in WP Prevention and Control Project

The goal of both clustering and classification methods is to distinguish data, but their implementation process is quite different. The classification problem is to select the best prediction model and classify and predict the new samples according to the existing label data:

3.1. Clustering Algorithm - K-Means Algorithm

The basic idea of K-Means algorithm is to randomly select the original cluster K in the data set, and then determine the Euclidean distance between the data object and the cluster. Finally, the data object in each cluster is calculated as the new cluster center, and then the next iteration is performed

until it changes.

In space, the Euclidean distance from the data object to the cluster center is calculated as follows:

$$d(X, C_n) = \sqrt{\sum_{m=1}^{j} (X_m - C_{nm})^2} \tag{1}$$

Among them, X is the data object; C_n is the n-th cluster center; M is the dimension of the data object.

In all data sets, the formula for calculating the sum of squares of errors is as follows:

$$SSE = \sum_{i=1}^{k} \sum_{X \in C_n} |d(X, C_n)|^2 \tag{2}$$

Among them, the size of the sum of squares of errors indicates the quality of clustering results; k is the number of clusters.

3.2. Classification Algorithm - Naive Bayesian Algorithm

Characteristic condition assumption: assuming that there is no correlation between the characteristics of the control project, a sewage sample training data set is given. In this data set, each sampling x contains an i dimension characteristic, that is to say:

$$x = (x_1, x_2, ..., x_i) \tag{3}$$

The class tag set contains several categories, namely:

$$y = (y_1, y_2, ..., y_i) \tag{4}$$

In a new sample x, people can determine which kind of sign it belongs to, and then determine the possibility that x belongs to class y_k according to Bayesian law.

$$P(y_k \mid x) = \frac{P(x \mid y_k) \times P(y_k)}{\sum_{k} P(x \mid y_k) \times P(y_k)} \tag{5}$$

The category with the largest posterior probability is recorded as the prediction category.

The simple Bayesian algorithm gives an independent conditional probability distribution. In short, it assumes that the attributes of each dimension are independent.

$$P(x \mid y_k) = P(x_1, x_2, ..., x_i \mid y_k) \prod_{n=1}^{i} P(x_n \mid y_k) \tag{6}$$

4. In-depth Exploration of Regional Water Body Prevention and Control Project

The pollution of water resources is a problem that is attached great importance to all over the world. Therefore, water prevention and control engineering workers must clearly understand its hazards and take targeted measures to maximize the economic benefits of environmental engineering. Table 1 shows the comparison of the number of urban and rural WP prevention projects in the selected areas:

Table 1. Quantity comparison of urban and rural water pollution prevention and control projects

	City(dwell)	Village(dwell)
2018	321	57
2019	345	69
2020	367	81
2021	381	78
2022	372	73

It can be seen from Table 1 that in the period from 2018 to 2022, the number of WP prevention and control projects in cities was far more than that in rural areas. Among them, the number of WP prevention and control projects in cities would increase from 2018 to 2021. However, the number of WP prevention and control projects in rural areas has increased from 2018 to 2020, which indicated that WP prevention and control work in rural areas was generally easier to carry out than in cities. Compared with the previous year, the number of urban WP prevention and control projects increased by 7.5% in 2019, 6.5% in 2020, 3.8% in 2021 and 2.4% in 2022; compared with the previous year, the number of rural WP prevention and control projects increased by 21.1% in 2019, 17.4% in 2020, 3.7% in 2021 and 6.4% in 2022.

Based on the analysis of the WP prevention and control projects to be optimized in two regions, the reference values of the natural and social conditions in the regions were compared first, and then the second experiment was carried out in one region with poor conditions according to its specific performance. Then, the K-Means algorithm and the naive Bayes algorithm were used to construct the WP prevention and control project in the selected area, and the prevention and control effects of the project before and after the use were counted.

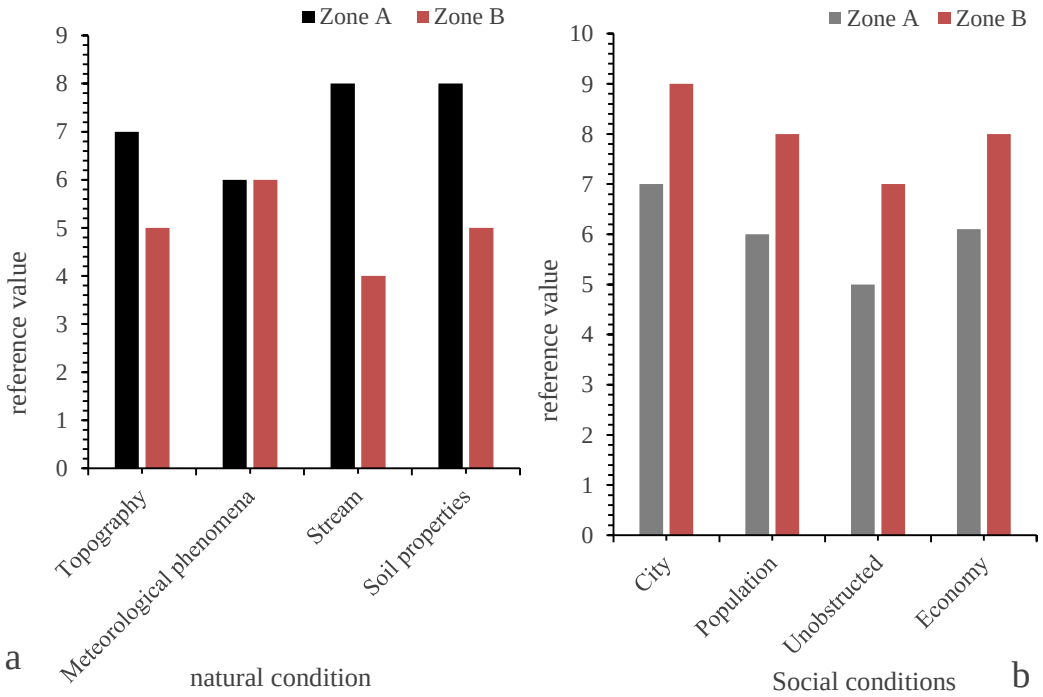

Figure 3a. Natural conditions for the 2 regions

Figure 3b. Social conditions in the 2 regions

Figure 3. Comparison of the natural and social conditions in the 2 regions

Figure 3a shows the reference values of four natural conditions in Regions A and B, and Figure 3b shows the reference values of four social conditions in Regions A and B. It can be seen from Figure 3 that the reference values of the other three natural conditions in Region B were lower than those in Region A, except for the reference values of meteorological conditions. It can be seen that the original conditions in this area were poor, which also determined the difficulty of carrying out WP prevention and control projects. In terms of social conditions, the urban, population, traffic and comprehensive economic conditions of Region B were higher than those of Region A. This region with poor natural conditions but extremely developed social conditions also determined the difficulty of its prevention and control project. Therefore, in order to make the research in this paper not lack universality, Region B was selected as the main body of the following experiments, and the method was applied to its local WP prevention and control project, as shown in Figure 4:

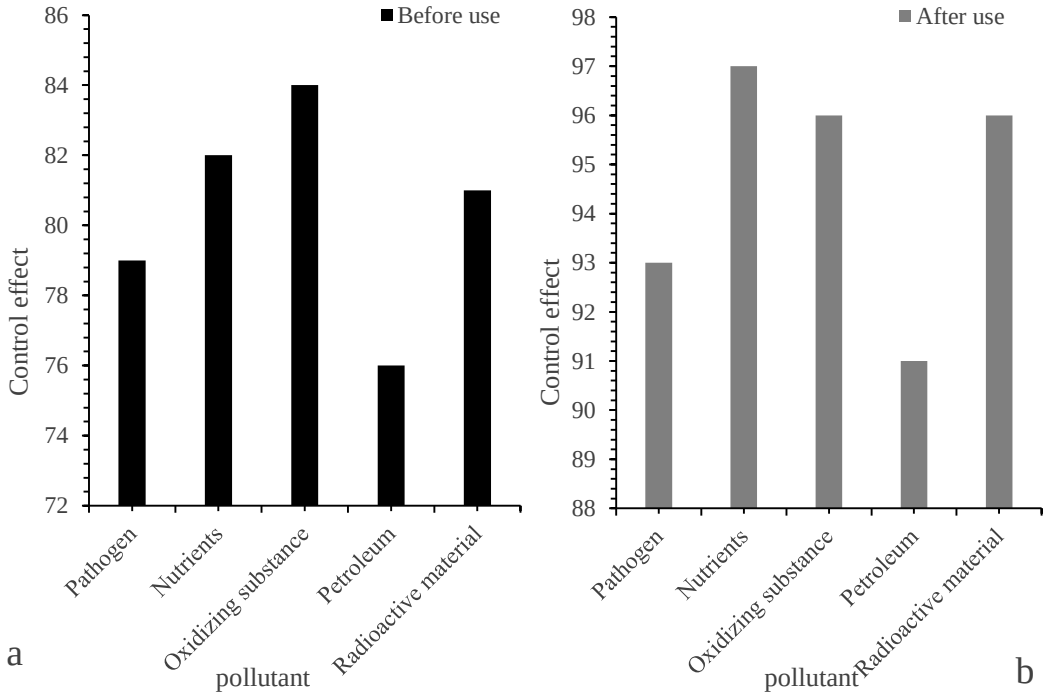

Figure 4a. Prevention and treatment effect before use

Figure 4b. Prevention and treatment effect after use

Figure 4. Prevention and treatment effect before and after use

Figure 4a shows the prevention and control effect of the WP prevention and control project in this area on water pollutants before use, and Figure 4b shows the prevention and control effect of the WP prevention and control project in this area on water pollutants after use. It can be seen from Figure 4 that the prevention and control effect of pathogens, nutrients, oxidants, oil and radioactive substances before use was worse than that after use. The WP prevention and control project based on clustering and classification algorithm can accurately detect the samples in WP, so it can be inferred that this kind of algorithm can be effectively applied to water environmental engineering.

5. Conclusion

Zero discharge of water pollutants is difficult or almost impossible, but sewage must be properly treated to ensure compliance with national or local pre-discharge standards, and the relationship

between industrial wastewater and urban wastewater needs to be reasonably emphasized. Based on the functions of the current water environment and the needs of economic and social development, the water quality grouping based on the surface water quality standards is the basis of protecting water sources and controlling WP. Therefore, to establish a scientific and intelligent WP prevention engineering network in urban buildings, more sewage treatment centers need to be built to reduce WP.

Funding

This article is not supported by any foundation.

Data Availability

Data sharing is not applicable to this article as no new data were created or analysed in this study.

Conflict of Interest

The author states that this article has no conflict of interest.

References

[1] Yubao Wang. Chinese industrial water pollution and the prevention trends: An assessment based on environmental complaint reporting system (ECRS). Alexandria Engineering Journal. (2021) 60(6): 5803-5812. https://doi.org/10.1016/j.aej.2021.04.015

[2] Mingjing He. Waste-derived biochar for water pollution control and sustainable development. Nature Reviews Earth & Environment. (2022) 3(7): 444-460. https://doi.org/10.1038/s43017-022-00306-8

[3] Long Yan. Comprehensive risk assessment of algae and shellfish in the middle route of South-to-North Water Diversion Project. Environmental Science and Pollution Research. (2022) 29(52): 79320-79330. https://doi.org/10.1007/s11356-022-21210-0

[4] Wei Hou. Occurrence and distribution of antibiotic resistance genes in lakes and reservoirs from water-receiving area of Eastern Route of the South-to-North Water Diversion Project, Northern China. Water Supply. (2020) 20(8): 3029-3037. https://doi.org/10.2166/ws.2020.190

[5] Lakshmikantha Varsha. IoT based smart water quality monitoring system. Global Transitions Proceedings. (2021) 2(2): 181-186. https://doi.org/10.1016/j.gltp.2021.08.062

[6] Xian Liu. Data-driven machine learning in environmental pollution: Gains and problems. Environmental science & technology. (2022) 56(4): 2124-2133. https://doi.org/10.1021/acs.est.1c06157

[7] Qingxuan Meng, Qingxuan, Jianzhuo Yan. A data cleaning method for water quality based on improved hierarchical clustering algorithm. International Journal of Simulation and Process Modelling. (2019) 14(5): 442-451. https://doi.org/10.1504/IJSPM.2019.104120

[8] Kumar K, Kishore, Kannan E. Detection of rice plant disease using AdaBoostSVM classifier. Agronomy journal (2022) 114(4): 2213-2229. https://doi.org/10.1002/agj2.21070

[9] Lodhi Pooja, Omji Mishra, Gagandeep Kaur. WQVP: An API enabled Open Data Machine Learning based Solution for Water Quality Visualization and Prediction. Journal of Telecommunication, Electronic and Computer Engineering (JTEC). (2018) 10(2): 61-72.

[10] Rastogi Rohit. Critical Analysis of Air Pollution and AQI and Its Bad Effects on Human Health: AI-and DL-Based Approach for Sustainable Development. International Journal of

Social Ecology and Sustainable Development (IJSESD). (2022) 13(1): 1-19. https://doi.org/10.4018/IJSESD.298334

[11] Jianhua Wu. Statistical and multivariate statistical techniques to trace the sources and affecting factors of groundwater pollution in a rapidly growing city on the Chinese Loess Plateau. Human and Ecological Risk Assessment: An International Journal. (2020) 26(6): 1603-1621. https://doi.org/10.1080/10807039.2019.1594156

[12] Nguyen Hiep Duc. Spatio-temporal pattern of water quality in the Saigon-Dong Nai river system due to waste water pollution sources. International Journal of River Basin Management. (2021) 19(2): 221-243. https://doi.org/10.1080/15715124.2019.1700513

[13] Mengzhi Ji. Bacteriophages in water pollution control: Advantages and limitations. Frontiers of Environmental Science & Engineering. (2021) 15(12): 1-15. https://doi.org/10.1007/s11783-020-1378-y

[14] Mingjing He. Waste-derived biochar for water pollution control and sustainable development. Nature Reviews Earth & Environment. (2022) 3(7): 444-460. https://doi.org/10.1038/s43017-022-00306-8

[15] Ahmed Shahid, Saba Ismail. Water pollution and its sources, effects & management: a case study of Delhi. Shahid Ahmed and Saba Ismail (2018)'Water Pollution and its Sources, Effects & Management: A Case Study of Delhi'. International Journal of Current Advanced Research. (2018) 7(2): 10436-10442.