Scholar Publishing Group
Water Pollution Prevention and Control Project
https://doi.org/10.38007/WPPCP.2020.010405
ISSN 2790-2730 Vol. 1, Issue 4: 41-50

Data Mining Water Pollution Prevention Project

Diana Raufelder*

Univ Amsterdam, Amsterdam, Netherlands

**corresponding author*

Keywords: Data Mining, Water Pollution, Artificial Intelligence, Pollution Prevention and Control

Abstract: Water environment is an important basis for urban development, which directly affects all aspects of people's production and life. However, with the improvement of people's living standards and the acceleration of urbanization, the water crisis has also become one of the focuses of social attention. In recent years, water pollution incidents have occurred frequently, causing serious environmental pollution and ecological damage, and the environmental situation is not optimistic. In the face of this situation, the government attaches great importance to strengthening the treatment of water pollution. However, due to the lack of scientific and effective management methods, its effect is not ideal, and even has a growing trend, which affects the sustainable use of water resources and restricts economic development and social progress. Therefore, how to realize the coordination and unification between water quality safety and environmental protection has become an urgent problem to be solved in this era. Data mining is a new method based on this demand. It can provide people with relevant information accurately and timely to guide decision-making to a certain extent, thus improving the operation efficiency of water treatment system and reducing the loss caused by human factors, which has brought great convenience to human survival, life and other aspects. This paper analyzed and studied the current situation of urban water environment. Through data analysis, the deficiencies and causes were found, and corresponding countermeasures were proposed to prevent and control pollution, so as to promote the healthy and orderly development of the water environment. The original water pollution prevention and control system was compared with the optimized prevention and control model based on data mining theory. The results showed that the use of artificial intelligence technology to monitor and manage the water quality can not only improve the management efficiency but also ensure the accuracy of the water quality monitoring results. The treatment cost was also reduced by about 24.03%. It can be seen that data mining has certain advantages in solving urban water pollution problems, providing people with a new way of thinking and making people have more confidence in improving environmental quality.

1. Introduction

The deterioration of water environment not only directly threatens people's health and quality of life, but also brings a series of hazards to human society. In the process of economic development, water pollution prevention and control has been highly valued by governments at all levels. However, at present, the whole society is in the rapid development period of industrialization and urbanization, and the discharge of industrial wastewater is increasing year by year. In particular, some enterprises' sewage is discharged into rivers or lakes without treatment, which has seriously damaged the ecological balance of river basins and led to the continuous decline of water quality. The ecological environment has been seriously damaged, exacerbating the occurrence of disasters and accidents. Therefore, it is of great practical significance to strengthen the treatment of water pollution.

The prevention and control of water pollution is a continuous and comprehensive work. Many scholars have studied this. Sarker, Bijoyee analyzed the impact of surface water and groundwater pollution on urbanization and industrialization, and discussed the relationship between water pollution and urban development on a spatial scale, providing reference for the control and governance of water resources crisis [1]. Yan Yan carried out ecological risk assessment on surface water pollution. He analyzed the factors affecting water quality, and counted the concentration value and distribution of various pollutants in each water body [2]. Dutta Pijush adopted the liquid flow control system of optimized genetic algorithm to conduct dynamic simulation research on water pollution and water resource utilization rate, and modeled and calculated the relationship between various parameters and water quality changes, which could better reflect the actual situation [3].

Li, Peiyue studied the sustainable development of water resources. He analyzed the possible environmental problems in the development and utilization of groundwater resources, and put forward corresponding countermeasures from economic and social aspects [4]. Hashmi Shahwaiz Ahmed conducted a comprehensive analysis of water pollution prevention based on the Internet of Things and cloud computing, and provided technical support for urban water resources management through water quality monitoring, water environment early warning, water ecological restoration and other business modules [5]. Wang Yubao evaluated the status of industrial water pollution and its prevention and control trend, and analyzed the factors affecting the concentration and discharge of pollutants in industrial wastewater. He pointed out the problems and measures that should be paid attention to in industrial pollution control [6]. The prevention and control of water pollution has the characteristics of complexity, dynamics and diversity, and requires comprehensive use of various technical means to achieve good results.

Dai Yingjie studied the current situation of groundwater pollution and its treatment methods, and analyzed and evaluated the groundwater quality in different regions. He summarized several commonly used treatment methods that had great impact on the groundwater environment [7]. Xu Zuxin studied the key technology of urban river pollution control, and established the technical system of groundwater environmental investigation, monitoring and early warning by analyzing the distribution law of pollutant concentration in the underground aquifer and the interaction between it and surface water [8]. Guo Qiaozhen conducted research on surface water change detection and environmental assessment. Starting with dissolved oxygen, total nitrogen, ammonia nitrogen and other indicators of water body, he combined the water environment capacity model and water quality index method to establish a set of mathematical models applicable to the change of groundwater in urban water supply plants [9].

Data mining has always been a research hotspot, and many scholars have conducted in-depth research on it from different perspectives. He Xiaodong investigated the surface water pollution in

the middle of the Loess Plateau and proved that the eutrophication of water bodies caused by agricultural production activities and industrial pollution was an important factor restricting the sustainable development of the ecological environment in the basin [10]. Morin-Crini Nadia studied the process and mechanism of water pollution by emerging pollutants and proposed a new water quality model based on multi-scale mixed modeling technology, which effectively solved the problem that it is difficult to accurately predict the spatiotemporal change trend of water in traditional water environment simulation [11]. Suriadikusumah Abraham used the pollution index method to analyze the water quality of the river, and used three indicators of pollution source, pollutant concentration and water eutrophication degree to evaluate the water environment in the river [12]. At present, many methods have been applied to water environmental quality assessment and water pollution prevention under different conditions. However, most of them lack pertinence and cannot meet the actual needs.

In order to solve the problems of monitoring and early warning in water pollution control and provide accurate and reliable basis and scientific and effective decision support for water pollution control, this paper proposed a comprehensive evaluation model of water environment quality based on data mining theory. This method was applied to the field of water pollution prevention and control, so as to reduce pollution sources and pollution levels, and achieve sustainable economic development. Compared with the neural network used in the prediction and analysis of environmental pollution in the past, the system has the characteristics of simple structure and easy programming, which is a hot topic in the research of water quality automatic monitoring technology at present.

2. Water Pollution Causes and Prevention Measures

2.1. Reasons for Water Pollution

Water pollution is a very serious environmental problem in the world today, especially the increasingly serious pollution of the chemical industry represented by oil, which has brought serious environmental pollution and ecological damage to the world. Therefore, water security has become one of the topics of common concern for all mankind [13]. There are various reasons for water pollution. In the past few decades, people have also conducted a lot of research and evaluation on water quality, and achieved many results. However, some new situations and new problems have also emerged, as shown in Figure 1.

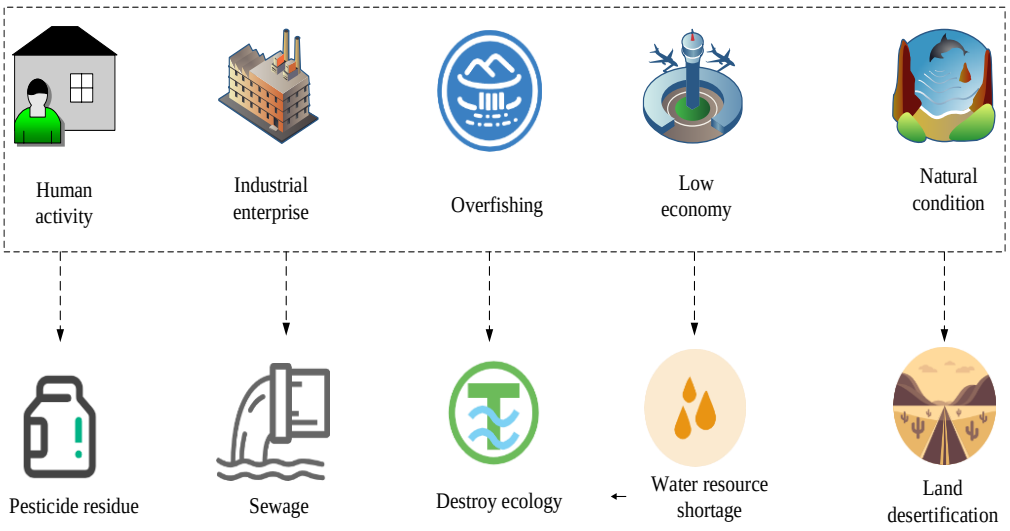


Figure 1. Causes of water pollution

The first is water eutrophication caused by human activities. This is because people need to constantly increase food intake in the long-term life process, and the food itself produces many toxic substances, such as heavy metals, pesticide residues, organic substances, etc. In addition, various industrial wastewater, agricultural irrigation water and livestock manure also cause eutrophication of the water body, thus threatening the quality of the water environment.

The second is the sewage discharged by industrial enterprises. The wastewater contains pollutants, which reduces the dissolved oxygen in the water and also brings a certain degree of biodiversity loss. Especially when industrial sewage is directly discharged into rivers, lakes or reservoirs, some aquatic organisms die, and in serious cases, fish diseases, parasitic diseases and even zoonoses may be caused.

The third is overfishing. The number of benthic animals decreases or disappears, especially in summer, some fishery operators blindly expand the area of net cage fish culture in order to achieve high profits in pursuit of economic benefits. There are also some fishermen who seek profits by using the method of "feeding fish" to make profits, causing aquatic animals and plants to be injured and harmed to varying degrees, which has seriously damaged the ecological balance and threatened the human living environment.

The fourth is the low level of economic development and the irrational industrial structure. Especially due to the continuous improvement of social productivity, the intensity of resource development has become greater and greater, resulting in the shortage of water resources, the increasingly serious water pollution, and the negative impact on the ecological environment.

The fifth is the deterioration of natural hydrological conditions. In particular, serious water and soil loss, land desertification and other geological changes have intensified, resulting in many rivers being cut off or dried up, and lakes decreasing or even disappearing. A large amount of sediment deposition has caused floods and coastal erosion, damaged the natural environment, and posed a serious threat to the safety of people's lives and property.

To sum up, water pollution has not only brought serious environmental pollution accidents, but also caused different degrees of economic losses to people. Therefore, it is one of the urgent tasks to strengthen investment in water pollution prevention and control and improve public participation awareness.

2.2. Development and General Application of Data Mining Technology

With the rapid development of science and technology, more and more industries have begun to pay attention to and use the Internet to explore the rules of user behavior. Therefore, data mining has emerged as the times require, and has been rapidly growing, showing a good development trend. As an emerging means of information acquisition and analysis, this technology is gradually penetrating into all aspects of society, providing people with more accurate and reliable data analysis, especially in the operation process of modern enterprises, which has a very wide range of applications, as shown in Figure 2.

First of all, library management system is a field in which data mining technology is applied in depth. The system can effectively improve the quality and efficiency of library management. Moreover, because it can accurately locate the needs of readers, it helps to promote university libraries to better serve students.

Secondly, for government management departments, data mining can scientifically process all kinds of data they have, so as to achieve the goal of more serving the people's livelihood. It can also help the government formulate more reasonable and feasible policies and programs to promote the healthy, stable and sustainable development of the economy.

Finally, in terms of water resources and environmental protection, data mining can better adapt to the characteristics of rapid environmental change and the complex background of various types

and uneven distribution of pollutants. At the same time, it also has strong classification ability and generalization ability, and can make qualitative judgment and quantitative description of water quality, so as to effectively alleviate the pressure and pollution of urban water supply [14].

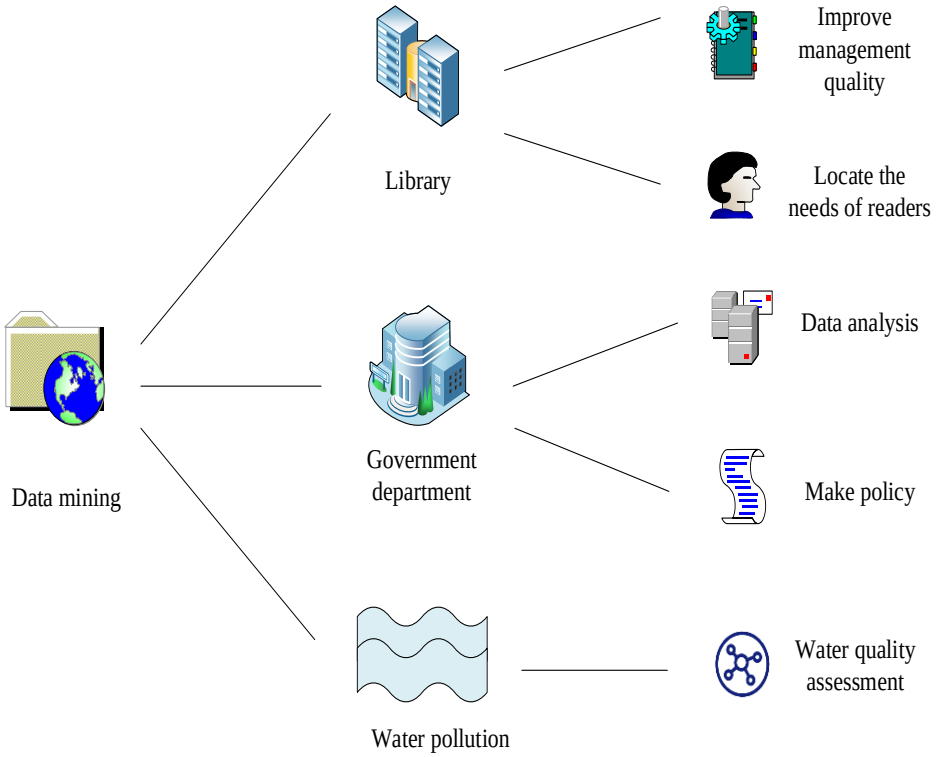


Figure 2. Development and general application of data mining technology

2.3. Water Pollution Control Model Based on Data Mining

With the improvement of people's living standards and consumption capacity, as well as the gradual formation and popularization of various environmental awareness, more and more people begin to pay attention to the solution of water pollution. Data mining is a new technology that extracts useful information through analyzing and processing a large amount of data, and then realizes the decision-making function of water resources management, which has a very broad application prospect in water conservancy projects. The water pollution prevention model based on this technology is shown in Figure 3.

The water pollution prevention and control model based on data mining mainly includes early warning and prediction module, risk assessment module and comprehensive evaluation module. Among them, early warning is a new method combining traditional hydrological forecasting methods with modern computer technology. It analyzes and processes the monitoring information to achieve real-time monitoring of water environment status and water resource management objectives. In the process of water pollution control, it can identify possible water quality problems in advance and take effective measures to control them in time, so as to reduce the incidence of pollution accidents and avoid major economic losses or casualties.

Risk assessment is a series of economic behaviors, such as making judgment on all kinds of flood accidents and their consequences in the basin, determining the degree of harm and causing losses. It is a complex and systematic system engineering, which requires a scientific and reasonable evaluation index system to provide basis for decision-making, so as to facilitate the analysis and prediction of different types of water environmental pollution events to achieve the

purpose of risk management.

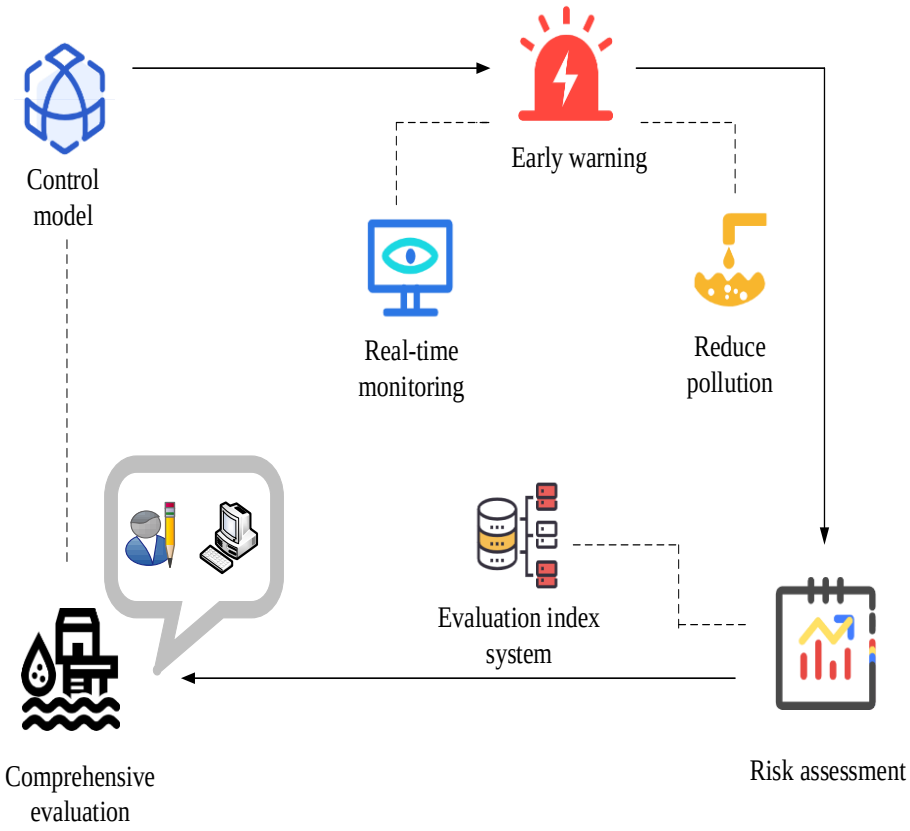


Figure 3. Water pollution control model based on data mining

The comprehensive evaluation includes not only the design of the evaluation index system, but also the establishment of corresponding mathematical models and the provision of quantitative indicators and calculation models, so as to better reflect the interrelationship and interaction between various subsystems, and finally achieve the goal of improving the work efficiency and saving the prevention and control costs.

The comprehensive application of these three modules can basically realize real-time online calculation and monitoring of pollutant emission concentration. At the same time, data fusion technology can also be used to correlate relevant indicators to form a complete indicator system. Then, the decision tree rule base is established based on rough set knowledge to improve the accuracy of water quality detection results, and the remediation of pollutants based on electrocatalysis can be realized on this basis [15].

3. Intelligent Technology of Data Mining

3.1. Clustering Mode

Clustering is a common method of data processing in data mining. In this kind of algorithm, the most important is the selection of core points. The calculation of average value based on data pair is the simplest choice method, and the specific expression is:

$$R_{ab} = \sqrt[i]{\sum_{x=1}^{N} |r_{ax} - r_{bx}|^{i}} \tag{1}$$

3.2. Regression Mode

Data fitting is based on the least square method, and its judgment formula is:

$$P = \sum_{x=1}^{N} [g(r_a) - j_a]^2 \tag{2}$$

The evaluation coefficient is used to compare the predicted sample results with the actual results. The calculation formula is:

$$q^2(j, \bar{j}) = 1 - \frac{\sum_{x=1}^{N} (r_x - \overline{r_x})^2}{\sum_{x=1}^{N} (j_x - \overline{j_x})^2} \tag{3}$$

In the formula, $\overline{j_x}$ is the average value of the calculation results, and N is the number of samples. $\overline{r_x}$ represents the prediction result.

4. Water Pollution Prevention Results Based on Data Mining

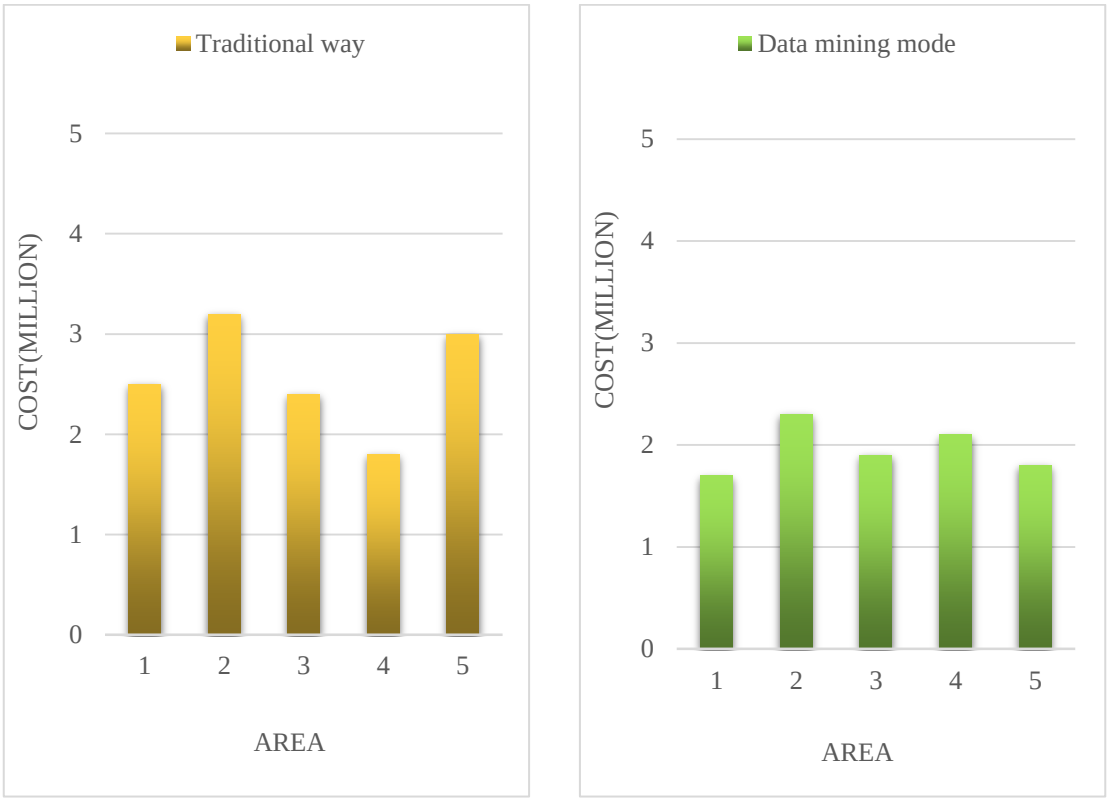
Taking a certain region as an example, five regions were randomly selected, and the suspended solids, chemical oxygen demand (COD), aerobic organic matter and pH value were taken as the evaluation indicators. There was serious water pollution in this region. Among them, the concentration of aerobic organics was expressed by biochemical oxygen demand 5 (BOD5) in five days, and whether there is water pollution in these areas under investigation. The results are shown in Table 1.

Table 1. Evaluation indexes of water pollution degree

	Area 1	Area 2	Area 3	Area 4	Area 5
PH	5.2	9	5.8	9.7	8.6
BOD5	120	104	189	160	141
Suspended matter	210	300	246	297	282
COD	240	210	371	330	286

As shown in Table 1, the PH values of these five areas were less than 6 or more than 8, showing a certain degree of weak acidity or weak alkalinity, and the suspended solids and COD content were not low. It indicates that the water quality in these areas is poor and may contain heavy metal ions. Therefore, it is necessary to strengthen the monitoring and prevention of water pollution in these areas to reduce waste on the premise of reducing the degree of water pollution, so that water conservancy projects can be carried out more scientifically and reasonably.

It can be seen from Table 1 that there are serious water pollution phenomena in these five regions. The traditional treatment methods and the water pollution monitoring and early warning model based on data mining were used to guide the water pollution management and prevention work in this region, and the treatment costs of the two methods were compared. The results are shown in Figure 4.



a. Traditional governance costs *b. Governance costs based on data mining*

Figure 4. Comparison of governance cost between the two methods

It can be seen from Figure 4 that Figure a shows the governance cost of traditional methods, and Figure b shows the governance cost based on data mining mode. It is obvious that the data in Figure a is larger than that in Figure b, indicating that the cost of traditional governance may be higher. According to the analysis, the cost of Figure a was controlled between 1 million and 4 million. Among them, area 2 had the most serious pollution and the highest cost, which has exceeded 3 million. Area 4 had the lowest cost, less than 2 million. It shows that the traditional governance mode is greatly affected by regional environmental changes and cannot achieve precise management and effective control. Looking back at Figure b, the cost data was basically distributed between 1-3 million, most of which fluctuated around 2 million. It shows that this treatment mode is relatively stable, and has a certain scale and sustainability, which can effectively reduce the degree of pollution, and improve the level of environmental quality and the quality of life of residents. After calculation, the average cost of the traditional governance method was about 2.58 million, and the average cost of the governance scheme based on data mining was about 1.96 million. The cost of water pollution treatment after the introduction of data mining technology was about 24.03% lower than that of traditional treatment. Therefore, data mining can be used for water environment management to reduce costs and environmental pollution, and achieve the goal of sustainable development.

5. Conclusion

The treatment of water pollution has always been the focus of social attention, and the automatic monitoring of water quality as an effective means can provide strong support for the online monitoring of pollution sources and the comprehensive management of water environment. The water pollution control model based on data mining can realize the prediction and early warning of

pollutant concentration, and can improve the environmental quality to a certain extent and enhance the public awareness of environmental protection. This paper analyzed the various causes of water pollution, and established a water quality automatic monitoring network system model based on data mining theory. The complete process from data acquisition to data processing was proposed. Combined with database technology, pattern recognition method and other key technologies, the effectiveness of the system for water quality monitoring and pollution control was verified through experiments, laying a foundation for better research in this field in the future.

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] Sarker, Bijoyee. "Surface and ground water pollution: causes and effects of urbanization and industrialization in South Asia." Scientific Review 7.3 (2020): 32-41.

[2] Yan, Yan. "Ecological risk assessment from the viewpoint of surface water pollution in Xiamen City, China." International Journal of Sustainable Development & World Ecology 25.5 (2018): 403-410.

[3] Dutta, Pijush, and Asok Kumar. "Modelling of liquid flow control system using optimized genetic algorithm." Statistics, Optimization & Information Computing 8.2 (2020): 565-582.

[4] Li, Peiyue, and Hui Qian. "Water resources research to support a sustainable China." International Journal of Water Resources Development 34.3 (2018): 327-336.

[5] Hashmi, Shahwaiz Ahmed, Chaudhry Fahad Ali, and Saima Zafar. "Internet of things and cloud computing‐based energy management system for demand side management in smart grid." International Journal of Energy Research 45.1 (2020): 1007-1022.

[6] Wang, Yubao. "Chinese industrial water pollution and the prevention trends: An assessment based on environmental complaint reporting system (ECRS)." Alexandria Engineering Journal 60.6 (2020): 5803-5812.

[7] Dai, Yingjie. "A review on pollution situation and treatment methods of tetracycline in groundwater." Separation science and technology 55.5 (2020): 1005-1021.

[8] Xu, Zuxin. "Urban river pollution control in developing countries." Nature Sustainability 2.3 (2019): 158-160.

[9] Guo, Qiaozhen. "An integrated study on change detection and environment evaluation of surface water." Applied Water Science 10.1 (2020): 1-15.

[10] He, Xiaodong, and Peiyue Li. "Surface water pollution in the middle Chinese Loess Plateau with special focus on hexavalent chromium (Cr6+): occurrence, sources and health risks." Exposure and Health 12.3 (2020): 385-401.

[11] Morin-Crini, Nadia. "Worldwide cases of water pollution by emerging contaminants: a review." Environmental Chemistry Letters 20.4 (2020): 2311-2338.

[12] Suriadikusumah, Abraham. "Analysis of the water quality at Cipeusing river, Indonesia using the pollution index method." Acta Ecologica Sinica 41.3 (2020): 177-182.

[13] Kosnik, Marissa B, Michael Zwicky Hauschild, and Peter Fantke. "Toward assessing absolute environmental sustainability of chemical pollution." Environmental Science & Technology 56.8 (2020): 4776-4787.

[14] LIANG, Xinqiang. "Promoting Construction of Zero-direct Discharge Engineering for Agricultural Wastewater from Paddy Fields in Yangtze River Delta Region." Bulletin of Chinese Academy of Sciences (Chinese Version) 36.7 (2020): 814-822.

[15] Singh, Nirala, and Bryan R. Goldsmith. "Role of electrocatalysis in the remediation of water pollutants." ACS Catalysis 10.5 (2020): 3365-3371.