Scholar Publishing Group
Water Pollution Prevention and Control Project
https://doi.org/10.38007/WPPCP.2020.010401
ISSN 2790-2730 Vol. 1, Issue 4: 1-10

# *Water Pollution Control Engineering Model of Waterworks Based on Clustering Algorithm and Machine Learning*

**Nicolas Bracikowski***

*Princess Nourah bint Abdulrahman University, Riyadh 11671, Saudi Arabia*

**corresponding author*

***Keywords:*** Clustering Algorithm, Machine Learning, Water Pollution of Waterworks, Pollution Prevention and Control

***Abstract:*** People can't live without water. As a kind of clean water and high-quality drinking water, tap water is more and more widely used in daily life, and is also favored by consumers. However, a large amount of industrial waste water and some toxic substances are produced during the water supply process, causing harm to human body. Therefore, it is necessary to take reasonable and effective measures to reduce the generation of these hazardous factors, so as to ensure the safety of water quality. Most of the traditional sewage treatment technologies rely on manual, mechanical or chemical methods for treatment, which consumes a lot of water resources. Moreover, with the continuous improvement of environmental protection requirements and the introduction of relevant laws and regulations, such sewage treatment methods are gradually replaced by green, intelligent and other new technologies. Advanced technologies such as clustering algorithm and machine learning (ML) provide new ideas to solve this problem. They have strong generalization ability and can automatically adjust their combination strategies according to the differences between different categories, thus achieving more efficient solution to pollution problems in complex environments. This paper introduced the main components of the water treatment system in the waterworks, and analyzed the current situation of water pollution treatment. At this stage, a series of problems in the water plant have been pointed out, and corresponding countermeasures have been put forward, such as optimizing the sewage treatment process, establishing a reasonable and effective scheduling mechanism, and strengthening the construction and management of sewage pipe network, which are of great significance to improve the water treatment rate and reduce energy consumption. The traditional governance methods were compared with the prevention model based on clustering algorithm and ML. The results showed that the optimized prevention and control model had strong pertinence, and could better achieve the water quality prediction function. The service quality has also improved by about 10.79%, which has achieved the goal of efficient and economic water pollution control, and has played a certain reference role for future practical work.

## 1. Introduction

The protection of urban drinking water sources and water quality safety are related to the health of the people, social harmony and stability and sustainable economic development. However, there is a serious shortage of water resources in most regions. However, the contradiction between supply and demand of water resources in these regions has become increasingly prominent due to the lagging economic development and the rapid increase in water demand caused by population growth. Therefore, the strengthening of the prevention and protection of urban drinking water source pollution has become a very important social responsibility.

A large amount of wastewater is produced after the operation of the water plant. Many scholars have studied this. Wu Gaojie analyzed the emergency response mechanism of different pollution sources, and believed that illegal discharge, sewage leakage and leakage during transportation were the main sources of chemical pollution, providing theoretical basis for reducing pollution, establishing and improving water pollution management [1]. Gibson Matthew encouraged enterprises to replace polluted watercourses, factory pollution and other media by specifying specific air emissions to treat polluted water bodies, so as to reduce pollutant emissions and improve energy efficiency [2]. Wang Shuming proposed several corresponding treatment processes for three types of water pollution sources, namely coal dust wastewater, domestic wastewater and oily wastewater, to achieve the water environmental protection objectives of river reach and wharf groundwater and other environmentally sensitive areas [3]. Singh Upma conducted a systematic study on the water pollution caused by industrial wastewater discharge, and proposed that biotechnology should be the main method, supplemented by chemical and physical measures. He combined a comprehensive treatment scheme combining multiple treatment methods to effectively solve the problems of high concentration and toxicity of organic substances in industrial wastewater [4]. Tang Yankui assessed the occurrence, monitoring, fate and risk of emerging pollutants in the water environment. He combined the traditional water environment capacity calculation method with new technologies to achieve the control of pollution sources [5]. Kanan Sofian used photocatalyst to degrade pesticides and major organic pollutants in water, and applied it to lakes and wetlands as a new ecological remediation method to achieve complete treatment in a relatively short time [6]. Water plant pollution is one of the important reasons for water shortage, resource waste and serious environmental pollution, and comprehensive measures must be taken to prevent its harm.

Clustering algorithm and ML have always been the research focus of the industry, and many scholars have carried out relevant research in this field. Khan Ahmed Shabbir studied the causes of river water pollution and its potential risks to human health, and believed that the high concentration of heavy metal ions in river water and serious pollution were the main reasons for the deterioration of river water quality [7]. Islam J. B studied the pollution of minerals and anions in surface water and their deposition trend on the sediments of Tula River to analyze the impact of these pollutants on water quality, and proposed measures to improve groundwater quality [8]. Haghiabi Amir Hamzeh investigated the performance of artificial intelligence technology, including artificial neural network, grouping data processing method and support vector machine for predicting the water quality composition of river water. He evaluated the accuracy of the application model according to the error index and found that support vector machine was the most accurate model [9]. Azrour Mourade used the advantages of ML algorithm to develop a model that could predict water quality index and then predict water quality category. He obtained the main pollution indicators and environmental risks of different water bodies by analyzing the data, and provided decision support for the government to formulate water pollution prevention policies [10]. Muhammad Salisu Yusuf proposed a water quality classification model based on ML algorithm,

which could quickly and accurately analyze water samples in complex environments, and improve the accuracy and real-time of water quality identification [11]. Ahmed Mehreen studied the ML technology of water quality index analysis and water pollution classification, which could quickly and effectively predict the water pollution level, and has been widely used in water treatment, industrial wastewater treatment, environmental monitoring and other fields [12]. ML and clustering algorithm play an important role in the development of automatic water quality evaluation system, water quality online monitoring system and sewage treatment industry.

In order to solve the disadvantages of traditional water pollution control technology in practical application and improve the operation efficiency and water quality safety level of the water treatment plant, this paper proposed an online sewage monitoring system based on clustering algorithm and ML. The system has collected, analyzed and processed the data generated in the process of sewage treatment in real time, and uploaded the information to the server to realize the remote control and management of sewage treatment. Compared with other sewage treatment methods, the system has the characteristics of low cost, strong real-time, and can well adapt to the current sewage discharge requirements under the complex environment.

## 2. Water Pollution Control of Water Works

### 2.1. Composition of Water Plant Treatment System

Rainwater, snow water, mineral water and other natural water can be used as domestic water or production water after certain process treatment and multistage purification. The water plant is a water intake facility that pretreates the natural water to make it reach a certain standard. Tap water supply is an important social public welfare undertaking, which is not only related to the immediate interests of the people, but also closely related to the development of the national economy. Therefore, management must be strengthened. The general composition of the treatment system of the waterworks is shown in Figure 1.

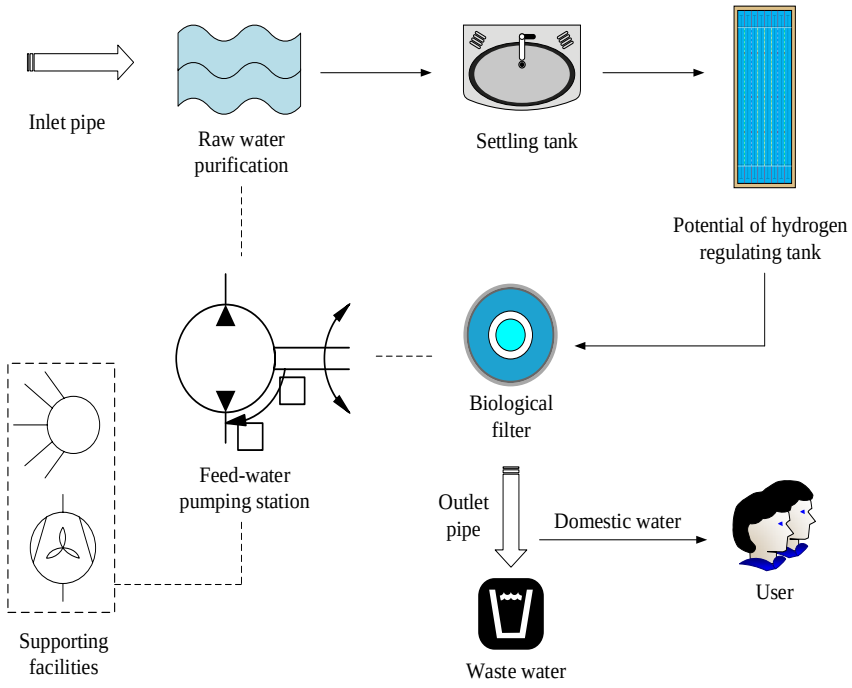



*Figure 1. Composition of waterworks treatment system*

The treatment system of the water plant is mainly composed of three parts. The first is sewage treatment device, including raw water purification device, sedimentation tank, pH adjustment tank and biological filter. This part mainly carries out advanced treatment of sewage and delivers the treatment results to downstream users through the corresponding pipe network. Among them, the sedimentation tank and acid-base regulating tank are secondary filtration devices used to remove suspended solids and a small amount of organic matter in the sewage. The biological filter is a three-stage filter, which is used to remove some organic substances in the water, so that they can be degraded to a certain extent or discharged completely up to the standard. The second is the water supply equipment and drainage pipeline system, which is composed of water inlet pipe and water outlet pipe. The water inlet end is connected to the water plant, and the water outlet at the water outlet end is connected to the drainage pipe respectively, thus forming a complete water supply and drainage system, ensuring the water supply capacity required for normal production and operation activities. The third is the water supply pump station and other supporting facilities, whose function is to provide water source and adjust the water volume according to the demand to meet the water requirements for different purposes.

With the aggravation of global warming, climate change and environmental pollution, the problem of water pollution has become increasingly prominent. In order to effectively control pollution, it is necessary to vigorously develop green circular economy and improve resource utilization efficiency and economic efficiency [13]. Generally, the reuse rate of urban domestic sewage is relatively high, which is because the membrane biological reactor (MBR) method is used for pretreatment of urban domestic sewage purification projects, and then sent to the reuse pipe network. After proper treatment, it reaches the drinking water standard, and finally the water body is discharged for surface water quality improvement or landscape environment construction, thus improving the utilization rate of water resources.

## 2.2. Current Situation of Water Pollution Control

In urban water supply, due to the influence of terrain conditions and groundwater resources, the quality of drinking water has been polluted to some extent, such as serious soil acidification and eutrophication of water around the drinking water source. Domestic sewage is discharged directly or mixed into surface water bodies, which causes great harm to the water environment. In particular, some industrial waste water contains a variety of toxic and harmful substances, which also leads to black and smelly water and heavy metal pollutants exceeding the standard, and even endangers human health [14]. Therefore, it is necessary to strengthen the construction of sewage treatment facilities and take effective measures to improve the quality of water environment to ensure the human survival environment and social sustainable development. The specific development status of water pollution control is shown in Figure 2.

At present, water pollution control technology includes three forms, namely, artificial wetland method, biological contact oxidation method and MBR method. Among them, the constructed wetland method is the most common and effective one. It can not only directly treat domestic sewage to zero discharge state, but also reduce the discharge of pollutants, so as to reduce the degree of water environment deterioration. However, this method requires a large amount of capital investment and cannot fundamentally change the urban water resources situation.

Biological contact oxidation method is mainly used for municipal sewage treatment. This method is to use the chemical reaction between microorganisms and sewage to degrade the wastewater after sewage purification. This method is simple, efficient and less investment. At present, it has a certain scale of operation experience. However, there are many problems in the actual operation, such as large energy consumption, complex operation, and easy to cause

secondary pollutants to exceed the standard.

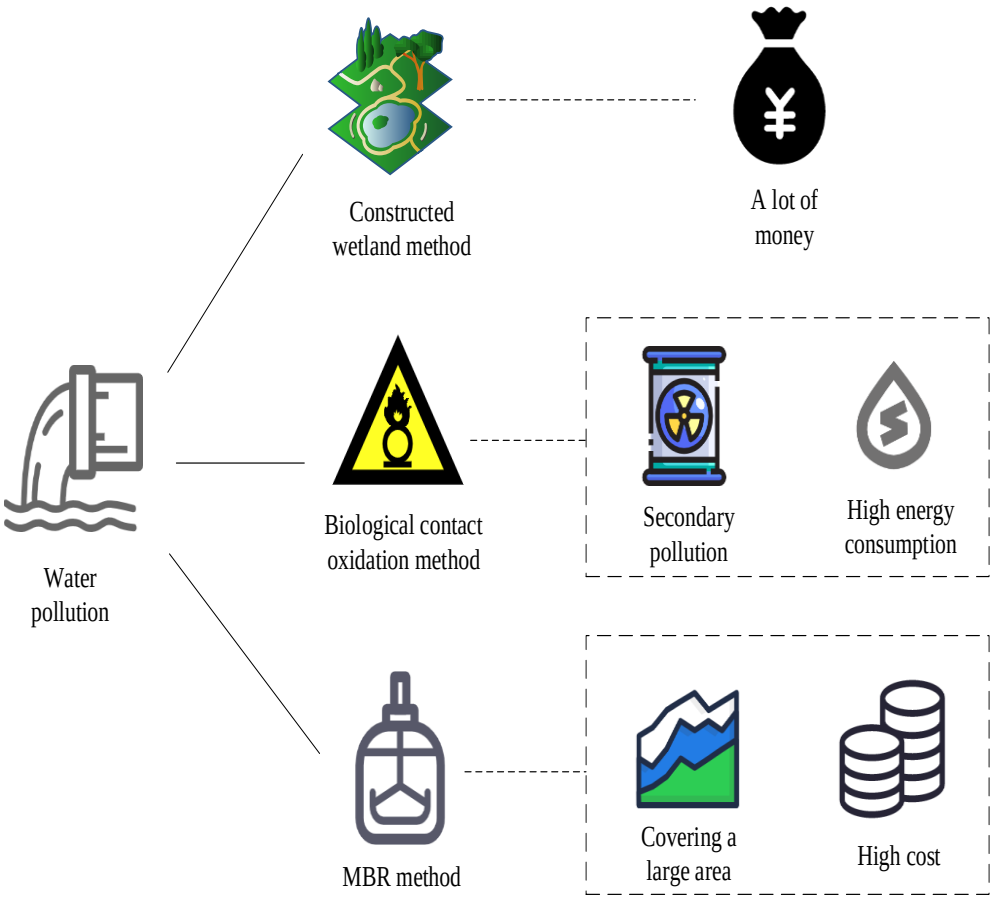



*Figure 2. Development status of water pollution control*

The MBR process uses domestic garbage collected manually or mechanically as raw material, and then dehydrates and dries it to make granular materials. Subsequently, it is sent to the treatment plant for further processing and molding, and finally transported to a water treatment process used by downstream users. It can realize the separation of sludge and industrial wastewater, so as to obtain a wastewater reuse treatment system that meets the requirements of water use. However, this technology has large floor area, high cost and high construction cost. Therefore, this process is not adopted in most regions at present.

## 2.3. Water Pollution Control Model of Waterworks

At present, there are many kinds of pollutants in the water plant with high concentration, and there are many impurities in the pipe network, including some substances harmful to human body. When they enter the water supply equipment, they may produce secondary pollution, resulting in threats to people's health [15]. In the process of industrial wastewater treatment, in addition to traditional physical methods, biological measures and the combined action of microorganisms and chemicals are also needed to effectively control the water quality index of wastewater. In addition, clustering algorithm and ML are also widely studied sewage treatment technologies. It can quickly and accurately predict the area with high degree of pollution and take it as the target function for the next step of treatment. These two treatment technologies are applied to the water pollution

treatment project of the waterworks, which can improve their operating efficiency, reduce operating costs and achieve the purpose of energy conservation and emission reduction. The governance model based on this construction is shown in Figure 3.

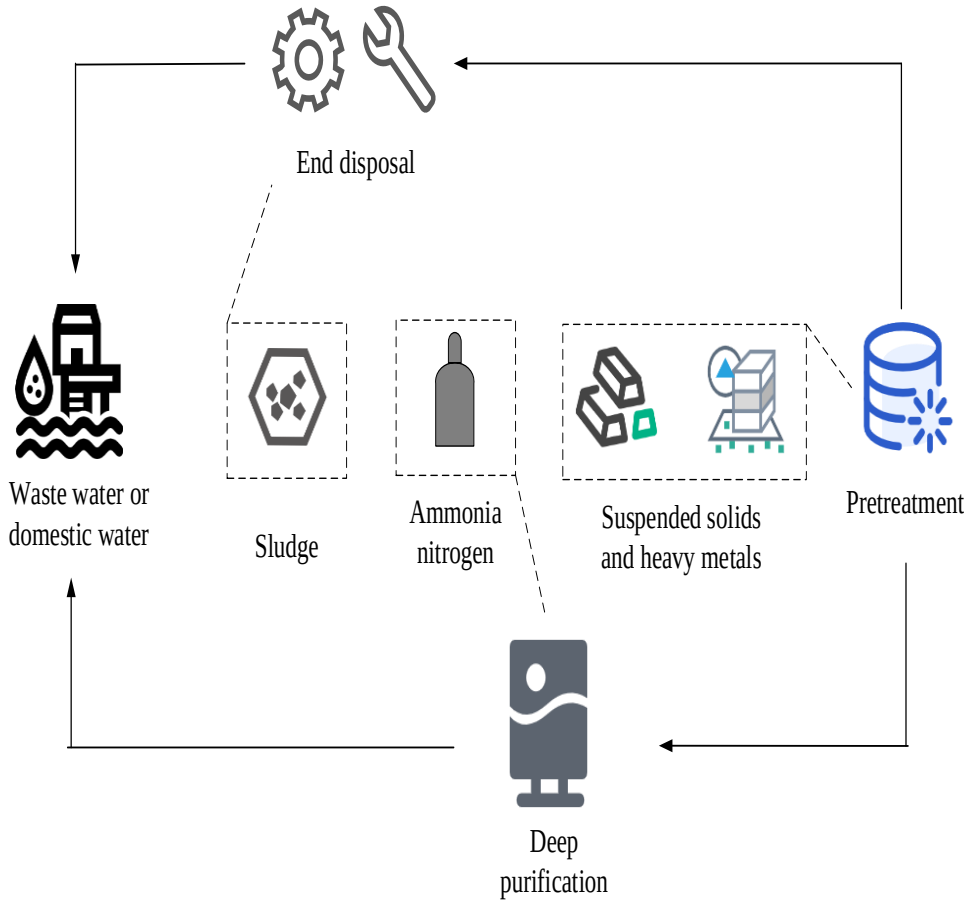



*Figure 3. Water pollution control model of waterworks*

The water pollution control model of waterworks based on ML and clustering algorithm includes three subsystems: pretreatment system, deep purification system and terminal disposal system. Among them, the pretreatment system is mainly used for preliminary screening of sewage to remove some suspended solids, organics, heavy metals and other impurities, so as to facilitate further filtration treatment in the later stage. The deep purification system is used to purify the remaining pollutants. By adding appropriate chemicals, most of the odorous substances in the water are removed, such as chloride ions, ammonia nitrogen, etc., which make the water clearer and improve the water treatment effect. The terminal treatment system is mainly used to provide the effluent quality requirements for subsequent processes, which can reduce the sludge amount generated in the sewage treatment process to a certain extent. Finally, the treated sewage is sent to the urban domestic sewage drainage pipe network or directly discharged into the river as industrial wastewater.

The water quality automatic clustering model constructed by ML method can establish a simple and reasonable clustering analysis model according to the information of sample set and the relationship between characteristic variables, and can complete the functions of similarity calculation, attribute sorting and membership function selection with traditional manual methods to

realize the combination of online monitoring, automatic calculation and remote monitoring. The traditional monitoring data collection system has been upgraded to the cloud platform and integrated with a variety of new technologies to form a complete set of automatic management methods. It not only has strong robustness, but also can automatically adjust the water treatment process parameters according to different environmental conditions to achieve the best removal effect and reduce the secondary pollution rate, which greatly improves the monitoring efficiency and service quality.

## 2.4. ML and Clustering Algorithm Technology

First of all, the clustering objective function is determined by integrating the intra-cluster variance relationship between clusters:

$$\alpha_{\max} = \max a_k = \max\left\{\sum_{x=1}^{N} \partial_{xk} \left| c_x - e_k \right|^2\right\} \tag{1}$$

Clustering minimization is performed. Based on the iterative formula, the relaxation formula is obtained:

$$\alpha_m = \sum_{k=1}^{M} m_k{}^n a_k = \sum_{k=1}^{M} m_k{}^n \sum_{k=1}^{M} \partial_{xk} \left| c_x - e_k \right|^2 \tag{2}$$

Among them, $m_k a_k \geq 0$, $\sum_{k=1}^{M} m_k = 1$.

In the process of solving, based on the Lagrange multiplier method, the following results are obtained:

$$m_k = \frac{a_k{}^{1/(1-n)}}{\sum_{k=1}^{M} a_k{}^{1/(1-n)}} \tag{3}$$

$$a_k = \sum_{x=1}^{M} \partial_{xk} \left| c_x - e_k \right|^2 \tag{4}$$

Among them, $0 \leq n \leq 1$, $1/(1-n) > 0$. The clustering process is regarded as the iterative updating process of the cluster and the cluster center, and the samples close to the center can be divided into clusters k. At this point, the larger the variance, the higher the weight.

## 3. Water Pollution Prevention and Control Results of Waterworks

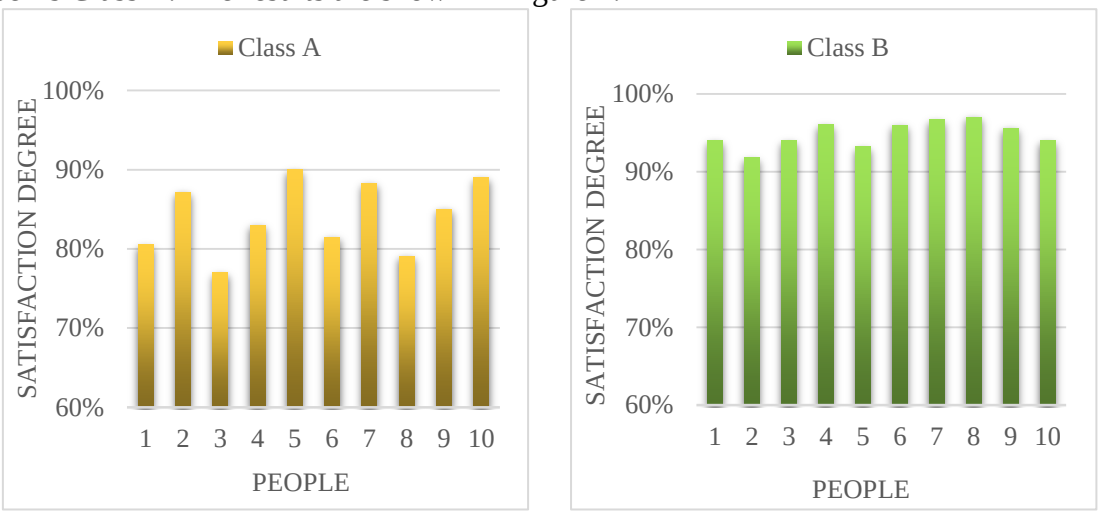
Five water plants were randomly selected from a certain area. The five plants all produced certain wastewater during the water treatment process. They were taken as experimental objects, and the traditional artificial treatment method and the control model based on ML clustering algorithm were adopted for simulation experiments. Among them, Factory 1 and Factory 2 adopted traditional methods, and Factories 3, 4 and 5 adopted prevention and control model based on ML clustering algorithm. The impact of different methods on sewage discharge and water quality parameters under the same water quality was analyzed. After treatment, the changes of chemical oxygen demand (COD), ammonia nitrogen and other pollutants in the waterworks were recorded, and the two removal effects were compared. The results are shown in Table 1.

*Table 1. Changes of COD, ammonia nitrogen and other pollutants*

| | COD | Ammonia nitrogen |
|---|---|---|
| Factory 1 | 23.6 | 3.6 |
| Factory 2 | 28.4 | 4.2 |
| Factory 3 | 19.1 | 1.4 |
| Factory 4 | 17.8 | 0.9 |
| Factory 5 | 20.5 | 1.1 |

As shown in Table 1, after treatment by some traditional biological and chemical means, the COD concentration of Factory 1 has dropped to 23.6mg/l, and that of Factory 2 has dropped to 28.6mg/l, which has reached the standard of reuse water and can meet production needs. The concentration of OCD in Factories 3, 4 and 5 decreased to 19.1 mg/l, 17.8 mg/l and 20.5 mg/l respectively. Compared with the other two factories, the governance mode based on ML and clustering algorithm prevention model has obvious advantages. To sum up, the treatment model using ML technology and clustering algorithm can better improve the water quality and inhibit the effect of sewage discharge on water pollution.

10 water quality experts were invited to evaluate the satisfaction of the two water pollution treatment methods. The traditional method is named Class A, and the control model proposed in this paper is Class B. The results are shown in Figure 4.




*a. Satisfaction with Class A methods* *b. Satisfaction with Class B methods*

*Figure 4. Comparison of satisfaction between two groups of governance styles*

As shown in Figure 4, Figure a shows the satisfaction of Class A mode and Figure b shows the satisfaction of Class B mode. It can be clearly seen that the data in Figure b were more concentrated and more than 90%, indicating that this type of governance has been unanimously agreed by experts. In Figure a, experts' scores were uneven, reaching a satisfaction level of about 90%. However, the minimum satisfaction rate was less than 80%, and the data fluctuation range was increased, with great uncertainty. After calculation, the average satisfaction of Class A governance was about 84.05%, while that of Class B was about 94.84%. Class B was about 10.79% higher than Class A. Therefore, water pollution control system based on clustering algorithm and ML model can more effectively classify water quality and realize intelligent management.

## 4. Conclusion

Water pollution prevention and control management of water works is a complex system engineering, involving water environment, water quality monitoring and other aspects. The combined application of clustering algorithm and ML can improve the efficiency of sewage treatment and reduce the labor intensity, so as to achieve the dual objectives of water resource conservation and environmental protection. This paper mainly studied how to use cluster analysis technology to classify and evaluate the pollution situation of waterworks and automatic early warning. ML model was introduced into the sewage treatment plant, and combined with fuzzy theory to realize the design and development of the sewage discharge prediction and automatic monitoring system, providing effective guidance for the sewage treatment of the waterworks.

## 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] Gaojie Wu. Water pollution management in China: recent incidents and proposed improvements. Water Science and Technology: Water Supply. (2018) 18(2): 603-611. https://doi.org/10.2166/ws.2017.139*

*[2] Gibson Matthew. Regulation-induced pollution substitution. Review of Economics and Statistics. (2019) 101(5): 827-840. https://doi.org/10.1162/rest_a_00797*

*[3] Shuming Wang, Lin Zhang. Water pollution in coal wharfs for coal loading and unloading in coal-fired power plants and its countermeasures. Journal of Coastal Research. (2020) 103(SI): 496-499. https://doi.org/10.2112/SI103-100.1*

*[4] Singh Upma. Water Pollution due to Discharge of Industrial Effluents with special reference to Uttar Pradesh, India-A review. International Archive of Applied Sciences and Technology. (2018) 9(4): 111-121.*

*[5] Yankui Tang, et al. Emerging pollutants in water environment: Occurrence, monitoring, fate, and risk assessment. Water Environment Research. (2019) 91(10): 984-991. https://doi.org/10.1002/wer.1163*

*[6] Kanan Sofian. Recent advances on TiO2-based photocatalysts toward the degradation of pesticides and major organic pollutants from water bodies. Catalysis Reviews. (2020) 62(1): 1-65. https://doi.org/10.1080/01614940.2019.1613323*

*[7] Khan Ahmed Shabbir. A review on current status of riverine pollution in India. Biosciences Biotechnology Research Asia. (2020) 18(1): 9-22. https://doi.org/10.13005/bbra/2893*

*[8] Islam J. B. Hydro-environmental pollution of Turag River in Bangladesh. Bangladesh Journal of Scientific and Industrial Research. (2018) 53(3): 161-168. https://doi.org/10.3329/bjsir.v53i3.38261*

*[9] Haghiabi Amir Hamzeh, Ali Heidar Nasrolahi, Abbas Parsaie. Water quality prediction using machine learning methods. Water Quality Research Journal. (2018) 53(1): 3-13. https://doi.org/10.2166/wqrj.2018.025*

*[10] Azrour Mourade. Machine learning algorithms for efficient water quality prediction. Modeling Earth Systems and Environment. (2020) 8(2): 2793-2801.*

*[11] Muhammad Salisu Yusuf. Classification model for water quality using machine learning techniques. International Journal of software engineering and its applications. (2018) 9(6): 45-52. https://doi.org/10.14257/ijseia.2015.9.6.05*

*[12] Ahmed Mehreen, Rafia Mumtaz, Syed Mohammad Hassan Zaidi. Analysis of water quality indices and machine learning techniques for rating water pollution: A case study of Rawal Dam, Pakistan. Water Supply. (2020) 21(6): 3225-3250.*

*[13] Zandalinas Sara I., Felix B. Fritschi, Ron Mittler. Global warming, climate change, and environmental pollution: recipe for a multifactorial stress combination disaster. Trends in Plant Science. (2020) 26(6): 588-599.*

*[14] Qiong Su. General equilibrium analysis of the cobenefits and trade-offs of carbon mitigation on local industrial water use and pollutants discharge in China. Environmental science & technology. (2019) 53(3): 1715-1724. https://doi.org/10.1021/acs.est.8b05763*

*[15] Xiaodong He, 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. (2020) 12(3): 385-401. https://doi.org/10.1007/s12403-020-00344-x*