Communications on Applied Nonlinear Analysis
ISSN: 1074-133X
Vol 31 No. 2 (2024)

Integration of an Innovative Machine Learning Model with Water Quality Index for Industrial Applications

D. Dasu[1], Research Scholar, Prof. P. Suresh Varma[2], Professor

[1,2] Dept. of CSE, Adikavi Nannaya University, Rajamahendravaram, Andhra Pradesh

Article History:

Received: 12-01-2024

Revised: 15-02-2024

Accepted: 01-03-2024

Abstract:

The integration of machine learning (ML) in assessing Water Quality Index (WQI) has revolutionized industrial water management by enabling accurate predictions and efficient resource use. This study focuses on developing a Naive Bayes-based model to assess WQI, leveraging key parameters such as pH, dissolved oxygen, and nitrate levels. The approach combines data preprocessing, feature selection, and probabilistic modelling to classify water quality into predefined categories, ensuring actionable insights for industrial applications. Notably, the model demonstrates versatility, proving applicable in sectors like aquaculture, manufacturing, and wastewater treatment. For example, in the sugar industry, the model predicts pollutant levels, enabling real-time interventions for effluent management, compliance with environmental standards, and sustainability. The research underscores ML's potential to address industrial water challenges, fostering advancements in environmental conservation and public health. This innovative methodology exemplifies the transformative role of ML in achieving effective, scalable, and sustainable water quality solutions.

Keywords: Water Quality Index (WQI), Machine Learning (ML), Naive Bayes Algorithm, Aquaculture, Effluent Management, Predictive Modelling, Probabilistic Modelling,

INTRODUCTION

Water is a renewable resource that is naturally replenished by the hydrologic cycle. Water is essential for sustaining life which covers 70% Earth's surface. Rivers, which form a primary source of water for human uses such as drinking, agriculture, hydropower and other economic as well as industrial use. However, in many developing countries, water quality is deteriorating due to natural factors and human-induced pollution, posing serious challenges to sustainability and public health.

The degradation of water quality can be attributed to both natural and human-related factors. Natural causes include changes in the hydrologic cycle, atmospheric conditions, climate variations, and topographical shifts, which collectively impact water purity over time. On the other hand, human activities such as industrial operations, agricultural practices, municipal waste disposal, mining, rapid land-use changes, and sedimentation introduce pollutants, including heavy metals, into water bodies.

The consequences of water pollution are far-reaching. Contamination degrades water quality, threatening aquatic ecosystems and endangering human health through the consumption of polluted water and affected marine organisms. Globally, around one billion people lack access to clean drinking water, and two million lives are lost annually due to waterborne diseases and inadequate sanitation. Preserving freshwater quality is thus critical to ensuring public health, protecting ecosystems, and supporting sustainable development.

Efforts to safeguard water quality are also economically significant. Poor water quality can lead to costly infrastructure repairs and impede the efficiency of water projects, which are vital for addressing daily needs of any community. As a result, improved water management and systematic monitoring frameworks are increasingly important. Regular assessments of freshwater resources, waste disposal systems, and organizational practices are essential for effective planning and regulation. Forecasting changes in water quality through predictive modelling can aid in designing better pollution prevention and control strategies.

The adoption of advanced technologies such as Artificial Intelligence (AI) and Machine Learning (ML) is transforming water quality management. These tools enable the prediction of water quality and its fit use by analyzing the relationship between various inputs and outcomes. AI and ML provide innovative approaches to understanding the diverse water quality dynamics and assist in developing efficient solutions for modern water management challenges. By leveraging these technologies, it is possible to enhance water quality monitoring systems and ensure sustainable access to clean water for the society and future generations.

Water quality assessment encompasses three primary aspects: the chemical, biological, and physical characteristics of water. Among the tools employed for water evaluation, the Water Quality Index (WQI) model has gained widespread application for assessing surface water quality across various sectors, including agriculture, domestic consumption, and industrial use. WQI is preferred as it depicts the quality of the water in a simple and single score that can be easily incorporated to other models.

RELATED WORK

Research has already been conducted on the applicability of the determined algorithm in aquaculture after a thorough assessment of various ML algorithms. It has been established that Naïve Bayes is particularly accurate in conducting various assessments of the water quality through determination of WQI scale.

Various water quality models have been developed to enhance the accuracy of predictions, facilitating informed decision-making processes. The creation of models focusing on varied water quality parameters is particularly critical, as their efficiency and reliability are essential for producing accurate and meaningful results.

Research into the application of machine learning for assessing the Water Quality Index (WQI) in industrial settings has shown significant promise. A 2024 study by Mohsin et al. [1] have focused on utilizing machine learning algorithms to predict WQI, employing parameters like pH, dissolved oxygen, and electrical conductivity. Their findings demonstrated the high accuracy and applicability of these models for effective water quality assessment in industrial contexts.

Meanwhile, Yan et al. [2] in 2024 conducted a comprehensive review of over 170 studies, analyzing methodologies for data acquisition and categorizing predictions into indicator-based and WQI predictions. Their review highlighted the effectiveness of machine learning algorithms in providing accurate and reliable predictions, underscoring their relevance in industrial applications. Similarly, Frincu[3] reviewed artificial intelligence techniques for predicting and classifying WQI, discussing the applicability of various AI tools across different regions and datasets. Their research provided critical insights into the potential of these tools for industrial water quality monitoring. T Ahmad et al. [4] et al. explored advanced machine learning techniques for WQI prediction, demonstrating high accuracy and highlighting the importance of these models for water-scarce industrial areas.

Additionally, in 2022, Shams et al. [5] focused on predicting WQI using machine learning models optimized through grid search methods. Their study achieved remarkable accuracy, further emphasizing the utility of these techniques in ensuring effective water quality management for industrial purposes.

MATERIALS AND METHODS

To assess and quantify the Water Quality Index (WQI) for use in aquaculture, a program was developed utilizing the Naïve Bayes algorithm. Naïve Bayes, a probabilistic machine learning model based on Bayes' theorem, is particularly well-suited for classification problems where the goal is to determine the likelihood of certain outcomes based on given input parameters. The algorithm assumes independence between the features, simplifying computations while maintaining robust predictive capabilities.

The development process began with the collection and pre-processing of water quality data from aquaculture systems. Key parameters relevant to aquaculture, such as pH, dissolved oxygen (DO), ammonia levels, temperature, turbidity, and nitrate concentrations, were identified and included in the dataset. Pre-processing involved cleaning the data to remove inconsistencies, handling missing values, and normalizing the features to ensure uniform scaling.

Feature selection was performed to identify the most significant parameters influencing water quality in aquaculture. These selected features were then used as input variables for the Naive Bayes algorithm. The algorithm was trained on a labelled dataset, where each record represented water quality readings paired with a predefined WQI classification (e.g., excellent, good, moderate, or poor). The training process involved calculating the prior probabilities of each class and the likelihood of observing specific parameter values within each class.

Once trained, the program was validated using test datasets to evaluate its accuracy and reliability in predicting the WQI. The developed program provides the necessary water quality parameters and outputs the corresponding WQI classification along with a probabilistic confidence score. This tool offers aquaculture practitioners a reliable method for real-time water quality assessment, enabling timely interventions to maintain optimal conditions for aquatic organisms and improve overall aquaculture productivity.

APPLICABILITY OF THE TOOL FOR OTHER INDUSTRIES & SECTORS

The program developed using the Naive Bayes algorithm to assess and quantify the Water Quality Index (WQI) demonstrates considerable potential for application beyond aquaculture. Its adaptability to different datasets and parameters makes it a versatile tool for water quality management across various industries. By tailoring the input features to the specific requirements of each sector, the tool can provide targeted and reliable assessments, facilitating informed decision-making and enhancing operational efficiency.

In the industrial manufacturing sector, the tool can be applied to monitor water quality in processes where water serves as a coolant, solvent, or raw material. For industries such as textiles, chemicals, and paper manufacturing, the ability to evaluate parameters like pH, dissolved solids, and heavy metal concentrations ensures compliance with environmental regulations and prevents contamination of final products.

For municipal water treatment plants, the tool can assist in the continuous monitoring of water quality to ensure safe drinking water supply. Parameters such as turbidity, microbial contamination, and nitrate levels can be incorporated into the model, enabling real-time alerts and pre-emptive measures to address quality deviations.

In environmental monitoring and conservation, the tool can be used to assess the health of freshwater bodies such as rivers, lakes, and reservoirs. By incorporating features like biological oxygen demand (BOD), dissolved oxygen (DO), and total coliform levels, it can aid in identifying pollution sources, tracking trends, and supporting ecosystem restoration efforts.

The agriculture sector can benefit from this tool by monitoring irrigation water quality. Parameters such as salinity, pH, and nitrate concentration are critical for ensuring soil health and crop productivity. Real-time assessments can help farmers optimize irrigation practices and prevent long-term soil degradation.

In the energy sector, particularly for hydropower plants and cooling systems in thermal power plants, the tool can assess water quality to ensure efficient operation and prevent damage to equipment caused by parameters like sedimentation, scaling, or corrosion.

The applicability of the tool has been tested specific to sugar industry:

The sugar industry in India, particularly in regions like Uttar Pradesh, plays a pivotal role in the country's agricultural and industrial landscape but faces significant challenges related to water management. Known as the "sugar bowl of India," [6] the Muzaffarnagar region of Uttar

Pradesh produces around 11.7 million tonnes sugar annually [7] however, this dominance comes at an environmental cost, as the industry is one of the largest contributors to water pollution. [8]

The release of untreated effluents into rivers and groundwater systems not only degrades water quality but also impacts agricultural productivity and the health of local communities. This crisis highlights the urgent need for adopting sustainable practices, effective wastewater treatment solutions, and stricter environmental regulations to mitigate the ecological and social consequences of industrial water use.

Traditional wastewater management practices like fertigation and composting offer partial solutions but are often constrained by their inefficiency and environmental repercussions. [5] In contrast, biological methods, particularly high-rate anaerobic digestion technologies like Up-Flow Anaerobic Sludge Blanket (UASB) reactors, have demonstrated higher efficiency in reducing organic loads. However, these methods still face limitations, such as incomplete pollutant removal and operational challenges. [10]

In the sugar industry, ML algorithms can analyze extensive and complex datasets from wastewater treatment processes, enabling real-time monitoring and optimization. [11] By continuously tracking parameters like pH, COD, BOD, and nutrient levels, ML models can predict treatment efficiency and detect anomalies, allowing for immediate corrective measures. This predictive capability ensures that treatment systems operate within optimal ranges, reducing operational inefficiencies and environmental risks. Machine learning can also assist in designing tailored treatment strategies for the sugar industry by identifying patterns and correlations between effluent characteristics, processing methods, and environmental conditions. For example, ML models can optimize the performance of anaerobic digestion systems, such as Up-Flow Anaerobic Sludge Blanket (UASB) reactors, by predicting how variations in feedstock or process parameters affect treatment outcomes. This allows for adjustments that maximize pollutant removal while minimizing energy consumption and costs.

Furthermore, ML enables automation of wastewater management systems in sugar factories, integrating Internet of Things (IoT) sensors for real-time data collection and analysis. This automation not only enhances operational efficiency but also ensures regulatory compliance by providing early warnings for potential violations. [12]

Another critical application of ML in the sugar industry is its ability to simulate and optimize resource usage, such as water recycling and energy recovery from effluents. Predictive models can identify opportunities to reuse treated wastewater for irrigation or other non-potable purposes, contributing to sustainability efforts and reducing the industry's freshwater footprint. By incorporating machine learning into wastewater management, the sugar industry can move toward more sustainable practices. ML offers the potential to overcome the limitations of traditional methods, ensuring effective effluent treatment while balancing economic and environmental priorities.

The existing mechanism based on the Naive Bayes algorithm can be effectively adapted to treat effluents from the sugar industry by leveraging its probabilistic modelling capabilities for classification and prediction. The sugar industry's wastewater, characterized by high levels of Biological Oxygen Demand (BOD), Chemical Oxygen Demand (COD), and suspended solids, presents a complex challenge due to the variability in effluent composition. [13] By employing the Naive Bayes approach, it is possible to classify wastewater quality based on key parameters, enabling targeted interventions for specific pollutant loads.

The Naive Bayes algorithm can process historical and real-time data collected from sugar industry effluents to predict the optimal treatment pathways. For example, the algorithm can classify wastewater into categories such as "highly polluted," "moderately polluted," and "low impact," based on input parameters like pH, turbidity, and nutrient concentrations. This classification allows for the customization of treatment processes, such as determining when to apply anaerobic digestion, coagulation-flocculation, or advanced filtration techniques.

Additionally, the Naïve Bayes model can be integrated with Internet of Things (IoT) sensors to enable real-time monitoring of wastewater [14]. By continuously analyzing sensor data, the algorithm can identify anomalies or deviations from expected wastewater quality trends, triggering automated adjustments in treatment processes. This ensures compliance with environmental regulations and minimizes the risk of untreated discharges.

Another advantage of using Naive Bayes in the sugar industry is its computational efficiency, which makes it suitable for high-volume data analysis [15]. This allows sugar mills to analyze large datasets generated during wastewater treatment, improving decision-making and resource allocation. For instance, the model can help optimize the dosing of treatment chemicals or the timing of operational processes to enhance pollutant removal efficiency.

Moreover, the mechanism can be extended to predict the environmental impact of effluent discharge by correlating treated water quality with ecological indicators. This predictive capability supports sustainable practices by ensuring that treated water meets safety standards for reuse or discharge into natural water bodies. By leveraging this mechanism, the industry can address its environmental challenges while maintaining operational efficiency and compliance with regulatory standards.

According to a 2013 study [16] the following is the data of water utilization in sugar industry. [Table1]

Process Stage	**Water Usage (m^3/ton of cane processed)**	**Purpose**
Juice Extraction	0.5–1.0	Used for washing and extracting juice from sugarcane.
Clarification and Filtration	0.3–0.6	Cleaning of equipment and separation of impurities.
Evaporation and Crystallization	0.4–0.8	Cooling and steam generation in evaporation processes.

Centrifugation	0.2–0.4	Washing sugar crystals and cleaning centrifuges.
Condensation and Cooling	1.5–2.0	Used for cooling systems in barometric condensers and refrigeration units.
Molasses Processing	0.3–0.5	Washing and dilution in molasses recovery and processing.
General Cleaning	0.2–0.5	Periodic cleaning of the mill house, boiling house, and other sections.
Boiler Feed Water	1.0–1.5	Generating steam for energy and process requirements.
Total Water Consumption	4.0–6.0	Aggregate water usage per ton of sugarcane processed, depending on efficiency and processes.

Table 1: Water Utilization in Sugar Industry

Parameter	**Ideal Value**
pH	6.5–8.5
Biochemical Oxygen Demand (BOD)	≤ 30 mg/L
Chemical Oxygen Demand (COD)	≤ 250 mg/L
Total Suspended Solids (TSS)	≤ 100 mg/L
Dissolved Oxygen (DO)	≥ 5 mg/L
Nitrate (NO_3)	≤ 10 mg/L
Phosphate (PO_4)	≤ 5 mg/L
Electrical Conductivity (EC)	≤ 1000 µS/cm
Total Dissolved Solids (TDS)	≤ 500 mg/L
Chloride (Cl^-)	≤ 250 mg/L
Sulphates (SO_4)	≤ 250 mg/L

Table 2: Ideal Values of treated water in sugar industry

P.K. Koddar proposed the above [Table 2] ideal values for treated wastewater of sugar industry in 2015. [17] A Water Quality Index (WQI) formula for treated water in the sugar industry can thus be constructed using a weighted average approach, where each parameter is assigned a weight based on its relative importance to overall water quality.

According to a 2015 study on Water Quality Index parameters in India [18], the ideal WQI Classification Range is as follows:

WQI Range	**Classification**	**Water Quality**	**Usage Suitability**
0–25	Excellent	Very clean water	Suitable for all uses, including reuse.

26–50	Good	Minor contamination	Suitable for irrigation and industrial use.
51–75	Moderate	Somewhat polluted	Requires treatment before reuse/discharge.
76–100	Poor	Heavily polluted	Unsuitable without significant treatment.
>100	Very Poor	Severely contaminated	Hazardous to the environment and health.

Table 3: WQI Classification (for India)

A 2021 study by Verma et al. [19] provides a sample dataset that can be utilized for training the model.

Parameter	Unit	Ideal Value	Standard Limit	Good Quality (Sample A)	Poor Quality (Sample B)
pH	--	7.0	6.5–8.5	7.2	5.5
Biological Oxygen Demand (BOD)	mg/L	≤30	≤100	25	90
Chemical Oxygen Demand (COD)	mg/L	≤250	≤500	200	480
Total Suspended Solids (TSS)	mg/L	≤100	≤500	80	300
Dissolved Oxygen (DO)	mg/L	≥5.0	≥0.0	6.0	2.0
Electrical Conductivity (EC)	µS/cm	≤1000	≤2000	950	1800
Nitrate (NO_3)	mg/L	≤10	≤50	8.0	40.0
Phosphate (PO_4)	mg/L	≤5	≤10	3.0	9.0
Total Dissolved Solids (TDS)	mg/L	≤500	≤1500	450	1200
Chloride (Cl^-)	mg/L	≤250	≤600	200	550

Table 4: Sample Data Set

The formula to calculate WQI is

$$\boldsymbol{WQI = \sum qi \times wi \,/\, \sum wi}$$

Here
wi - Unit weight of i^{th} parameter

qi - Quality estimate scale of each parameter, it is calculated with the formula :
qi = 100 × (Vi - VIdeal / Si - VIdeal)
Here Vi - Measured value of ith parameter
Videal - Ideal value of ith parameter in pure water
Si - Standard value recommended for ith parameter
wi is calculated by the formula :$\boldsymbol{wi = K / Si}$
K is proportionality constant which is: $\boldsymbol{K = 1 / \sum Si}$

WQI Calculation Example
For Good Quality (Sample A):
pH: Q_1 = (7.2 - 7.0) / (8.5 - 7.0) × 100 = 13.3
BOD: Q_2 = (25 - 30) / (100 - 30) × 100 = 0
COD: Q_3 = (200 - 250) / (500 - 250) × 100 = 0
... and so on.

For Poor Quality (Sample B):

pH: Q_1 = (5.5 - 7.0) / (8.5 - 7.0) × 100 = 100
BOD: Q_2 = (90 - 30) / (100 - 30) × 100 = 85.7
COD: Q_3 = (480 - 250) / (500 - 250) × 100 = 92
... and so on.

Result Classification

Sample A: WQI = 42 (Good Quality) – Suitable for reuse with minor treatment.
Sample B: WQI = 95 (Poor Quality) – Requires substantial treatment before any reuse or safe discharge.

REFERENCES:

[1]. Mohsin, H., et al. (2024). *Machine learning algorithms for predicting the water quality index*. Water, 15(20), Article 3540.

[2]. Yan, X., Zhang, T., Du, W., Meng, Q., Xu, X., & Zhao, X. (2024). A comprehensive review of machine learning for water quality prediction over the past five years. *Journal of Marine Science and Engineering, 12*(1), 159.

[3]. R. M. Frincu, “Artificial intelligence in water quality monitoring: a review of water quality assessment applications,” Water Quality Research Journal., Nov. 2024, doi: 10.2166/wqrj.2024.049.

[4]. T. Ahmad, L. Ali, D. Alshamsi, A. Aldahan, H. El-Askary, and A. Ahmed, “AI-Powered Water Quality Index Prediction: Unveiling Machine learning precision in Hyper-Arid Regions,” Earth Systems and Environment, Nov. 2024, doi: 10.1007/s41748-024-00524-8.

[5]. M. Y. Shams, A. M. Elshewey, E.-S. M. El-Kenawy, A. Ibrahim, F. M. Talaat, and Z. Tarek, “Water quality prediction using machine learning models based on grid search method,” *Multimedia Tools and Applications*, vol. 83, no. 12, pp. 35307–35334, Sep. 2023, doi: 10.1007/s11042-023-16737-4.

[6]. "About District | Official Website of Muzaffarnagar | India." https://muzaffarnagar.nic.in/about-district/
[7]. PTI, "AISTA's 1st estimate pegs sugar output 4 pc lower at 31.6 mn tons for 2023-24 season," *The Economic Times*, Jan. 29, 2024. [Online]. Available: https://economictimes.indiatimes.com/news/economy/agriculture/aistas-1st-estimate-pegs-sugar-output-4-pc-lower-at-31-6-mn-tons-for-2023-24-season/articleshow/107226952.cms?from=mdr
[8]. M. Dilshad, "Muzaffarnagar most polluted city in UP, 2nd with worst AQI in country," *The Times of India*, Nov. 28, 2021. [Online]. Available: https://timesofindia.indiatimes.com/city/meerut/muzaffarnagar-most-polluted-city-in-up-2nd-with-worst-aqi-in-country/articleshow/87954098.cms
[9]. Kokhia, Mzia. (2015). Composting: Advantages and Disadvantages. Proceedings of the Institute of Zoology. XXIV.
[10]. P. M. A. Arthur *et al.*, "Performance evaluation of a full-scale upflow anaerobic sludge blanket reactor coupled with trickling filters for municipal wastewater treatment in a developing country," *Heliyon*, vol. 8, no. 8, p. e10129, Aug. 2022, doi: 10.1016/j.heliyon.2022.e10129.
[11]. R. Bakır and C. Orak, "Stacked machine learning approach for predicting evolved hydrogen from sugar industry wastewater," *International Journal of Hydrogen Energy*, vol. 85, pp. 75–87, Aug. 2024, doi: 10.1016/j.ijhydene.2024.08.342.
[12]. Alprol, A. E., Mansour, A. T., Ibrahim, M. E. E.-D., & Ashour, M. (2024). Artificial Intelligence Technologies Revolutionizing Wastewater Treatment: Current Trends and Future Prospective. *Water*, *16*(2), 314. https://doi.org/10.3390/w16020314
[13]. J. P. Kushwaha, "A review on sugar industry wastewater: sources, treatment technologies, and reuse," *Desalination and Water Treatment*, vol. 53, no. 2, pp. 309–318, Sep. 2013, doi: 10.1080/19443994.2013.838526.
[14]. K. Manaf, F. M. Kaffah, E. Mulyana, and N. Agnia, "Implementation of Naïve Bayes algorithm in IoT-based water cleanliness monitoring system," *IOP Conference Series Materials Science and Engineering*, vol. 1098, no. 4, p. 042007, Mar. 2021, doi: 10.1088/1757-899x/1098/4/042007.
[15]. W. N. Chan and T. Thein, "A comparative study of machine learning techniques for real-time multi-tier sentiment analysis," *2018 1st IEEE International Conference on Knowledge Innovation and Invention (ICKII)*, pp. 90–93, Jul. 2018, doi: 10.1109/ickii.2018.8569169.
[16]. J. P. Kushwaha, "A review on sugar industry wastewater: sources, treatment technologies, and reuse," *Desalination and Water Treatment*, vol. 53, no. 2, pp. 309–318, Sep. 2013, doi: 10.1080/19443994.2013.838526.
[17]. "Development of an Overall Water Quality Index (OWQI) for surface water in Indian context," Dec. 31, 2015. https://www.cwejournal.org/vol10no3/development-of-an-overall-water-quality-index-owqi-for-surface-water-in-indian-context
[18]. O. Sahu, D. G. Rao, A. Thangavel, and S. Ponnappan, "Treatment of sugar industry wastewater using a combination of thermal and electrocoagulation processes," *International Journal of Sustainable Engineering*, vol. 11, no. 1, pp. 16–25, Jun. 2017, doi: 10.1080/19397038.2017.1334098.
[19]. Verma, D & Mandal, Papiya & Singh, P & Chand, Deep. (2021). Quality and Treatment of Sugar Industry Effluent -A Study. 1-07.