



Water Quality Monitoring using Machine Learning

- Sonal Kumari • Simran Sinha • Tannu Kumari
- Anshu

Received : November 2019

Accepted : March 2020

Corresponding Author : Anshu

Abstract : *Although water is the most important element present for the survival of mankind, there has been a tremendous increase in its contamination level due to rising population and urbanization.*

In our research paper we will focus on the monitoring of water using machine learning technology. Technology has led to the development of various tools which will create a better scope for future generations to make things smarter and faster. The main goal of the water monitoring system using machine learning is to enhance the prior technology that uses sensors.

Machine learning used along with sensors makes the water monitoring system more advanced as the water quality based on certain water parameters, like the chemical level and the contamination source, can be known within a few seconds by the cloud if it detects any illogical balance in water. Therefore, through proposed system we will try to know the contamination source in Yamuna river considering some water parameters as important factors. The main aim of the paper is to observe the role of machine learning in water quality monitoring which is more advanced than sensor technology used in traditional systems.

Keywords: *Machine Learning, Water Monitoring.*

Sonal Kumari

BCA - III, Session: 2017-2020,
Patna Women's College, Patna University, Patna,
Bihar, India

Simran Sinha

BCA - III, Session: 2017-2020,
Patna Women's College, Patna University, Patna,
Bihar, India

Tannu Kumari

BCA - III, Session: 2017-2020,
Patna Women's College, Patna University, Patna,
Bihar, India

Anshu

Assistant Professor, Department of Computer Science,
Patna Women's College, Bailey Road,
Patna-800 001, Bihar, India
E-mail : sayhi2anshu@gmail.com

Introduction :

Water is important for the survival although its contamination due to various human activities has been increasing at an alarming rate. Different water companies which supply water had taken many preventive measures to stop the growth of water contamination by using traditional method of using different water sensors to monitor relevant water quality which doesn't provide an automated system. It's here that machine learning comes into play. The word "Machine learning" was first given by Arthur Samuel in 1952 which means training the system to perform a

specific task without being explicitly programmed. Machine learning is acting as an innovation for modern era not only professionally or at industrial processes but also in our everyday living, for example, in medical diagnosis this technology is helpful in detecting various diseases.

Objectives :

- To find out different chemical levels dissolved in water.
- To find out future prospects and planning.

Hypothesis :

- Water Quality Monitoring using Machine Learning is a more advanced and better technology as compared to the traditional sensor using technology
- Traditional method using water monitoring system gives poor results as in sensor-based technique continuous observation of the water contamination level is not possible since it doesn't work with live datasets and source of contamination is difficult to find out (<https://www.usgs.gov/nwis-national-water-information-system>).

Different Types of water Monitoring system :

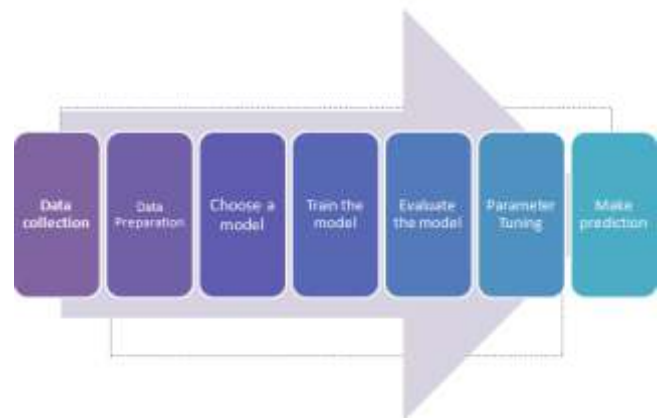
Hydraulic Information System : This type of water quality monitoring system measures water quality and management, i.e, measures , processes ,stores and also interlinks different environmental factors.

Surface water Monitoring System : This type of water monitoring systems measures water level and discharge rates. It determines the time, space of flow and water level throughout the river used in flood control.

Groundwater Monitoring System : This type of water monitoring system collects information related to spatial and temporal distribution of water level and hydraulic properties of underground layer of water bearing.

What is Machine Learning ?

Machine learning is the ability of the system which is trained in such a way that the machine learns automatically without being explicitly programmed.



Different steps involved in machine learning:

Data Collection-It involves the process of collection of data from various sources.

Data Preparation-It involves cleaning and transforming raw data prior to processing and analysis.

Choose a Model –It involves a process of selecting different models that generate predictions by finding patterns in the data.

Train the model-It involves different machine learning algorithm to be applied with training data

Evaluate the model-To find the best model that represents the particular data and how well the model will work in the future.

Parameter Tuning-It governs the training process itself.

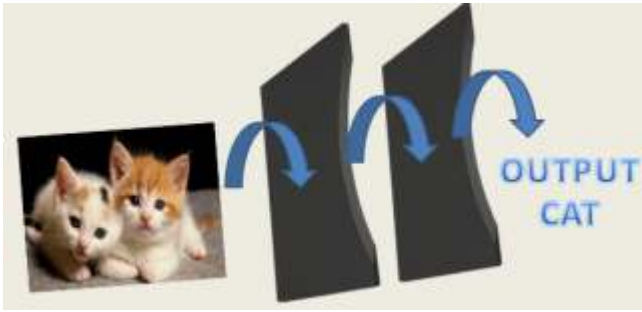
The parameters or the variables considered as parameter tuning does not directly relate to the training data also known as **hyperparameters**

Types of Machine Learning:-

1. Supervised Machine Learning
2. Unsupervised Machine Learning
3. Reinforcement Machine Learning

Supervised Machine Learning- It builds the model on the basis of input variable and output variable and trains the model to make predictions in which the new input data is given according to which the data for output variable is predicted.

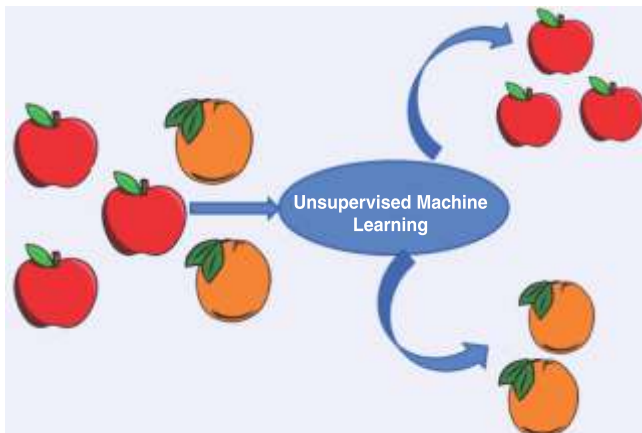
Supervised Machine Learning uses certain algorithm, such as, Support Vector Machines, classification, linear regression. The best example of Supervised Machine Learning is image classification in which image dataset is given and output is detected.



Unsupervised Machine Learning: In unsupervised machine learning the input data is given but the output data is not known. In order to learn about the data, the main goal of unsupervised machine learning is to know the hidden patterns in the given input data or the dataset to predict the output.

The best known algorithm for unsupervised machine learning are clustering, Association, K-means.

An example of unsupervised Machine Learning is shown below:



It discovers hidden structures and forms clusters.

Reinforcement Machine Learning: Reinforcement learning is a type of machine learning technology that allows an agent to interact with the environment and form its own actions and experiences. The main goal of reinforcement learning is to find an action model which would maximize the total reward of the agent.

Smart Water quality monitoring using IOT technology : Real time water quality observation is done using this technique (IOT) involving the steps of collecting and analysis of data of different locations.

The data here refers to different water parameters that we can consider to get a close indication of contamination of water.

Example of water parameters can be PH, temperature, Oxidation and reduction potential and many others.

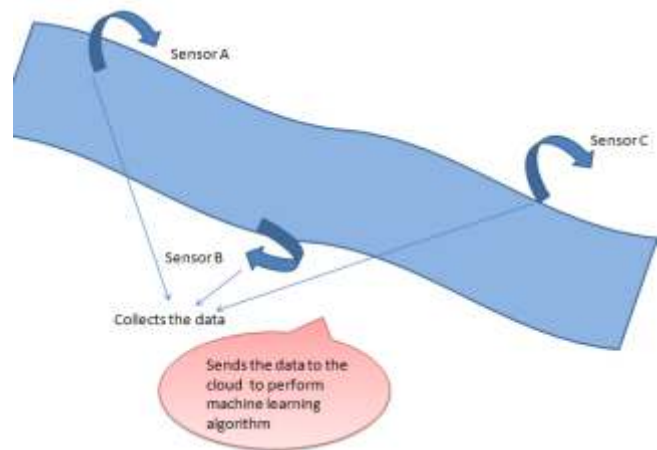
Water Quality Monitoring system applied on Machine Learning. Although the sensor technology is the root of the monitoring systems. Different appropriate sensors installed at proper locations send the data accordingly to the respective individuals or monitored systems. But the traditional sensor technology has certain limitations:

- Not working on real time data accurately can give inappropriate results.
- Automated Monitoring cannot be possible
- Contamination source cannot be known.

So, the tremendous growth in technology has given us many tools we can use to make our life easier and not considered as a threat.

Suppose if we want to check the quality of water first we have to take different samples of water from the source varying different locations and the next step is to take these samples to the laboratories. In laboratories based on certain parameters the quality of water is tested (Overview of Clean Water AI).

The systematic diagram of proposed Water monitoring system :



This proposed system deals with automatic monitoring on the real time data and reacts accordingly.

Considering 3 sensors— sensor A , sensor B, sensor C – dipped into Yamuna river at fixed distance, i.e, 100 metres. When water flows from sensor A to sensor B these wireless sensors simply collect the data and send to the cloud. Depending on the water parameters cloud collects the data and performs machine learning algorithm.

According to the research, the best suited machine learning algorithm for water quality monitoring are isolation forest and deep learning algorithm like auto encoder. In isolation forest, the word isolation means separating an instance from the rest of the instances and can be used in anomaly detection which is the identification of rare items which raise suspicions by differing from the majority of data.

Working of Proposed System of Water Monitoring

Step 1: Collection of data from respective sensors.

Step 2: The data collected dumped to the cloud.

Step 3: Automated machine learning is used for feature processing, feature selection, model selection and hypoparameter tuning.

Step 4: Since the algorithm performed should be able to handle large dataset, thousands of pipeline responsible to find the best solution for the particular problem, train the model and evaluate the performance using cross validation. Since the output data is not given, the prediction is done using unsupervised machine learning for this proposed water monitoring system the best suited algorithm according to the result providing higher accuracy is isolation forest algorithm.

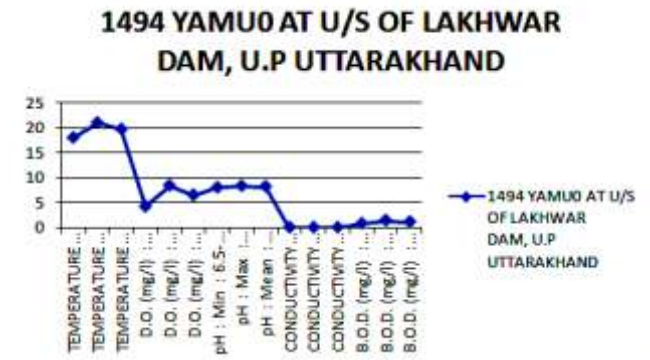
When data is dumped streaming data is important as it is related to real time processing of data.

When data is streamed in real time inside the system as when data is pushed into the target application using publish-subscribe approach API(Application Programming Interface)real time alert is created as data processing is done in real time. In stream data the approach of distributed computing is used the model is trained (Trained data)and the new set of data is given (test data)to predict the result.

Graph showing the Yamuna river temperature, PH, conductivity, B.O.D of Yamuna river in year 2013 in Uttarakhand.



Graph showing the Yamuna river temperature, PH,conductivity, B.O.D of Yamuna river in year 2014 in Uttarakhand



Reasons of water impurity

- Due to the increased population leading to rapid urban development
- Excessive use of water resources.
- Dumping of chemical waste by factories leading to the deterioration of the quality of water.
- Use of pesticides and insecticides in agriculture and domestic waste.
- Establishment of large industrial houses and increased number of industry.
- Improper sewage disposal.
- Throwing of plastic in the ocean resulting in contamination harming the aquatic ecosystem.

Methodology Used:

Area of study - Various rivers datasets, water monitoring companies, and environmental data on their official sites and from different water sources.

Sample size and sampling method – Samples will be collected through questionnaires and interview with students and research scholars of different age groups and also from those people who don't belong to any educational institutes.

Tools and technique – Interviews, questionnaires, magazines, journals, newspaper, websites, and articles will be used in collecting all the relevant information for the research .

Data analysis – Data will be of two types, first will be primary data which will be collected from questionnaires and interviews and the second will be the secondary data that will be collected from different sources like different articles, journals, websites, and newspaper.

Limitations of water quality monitoring using machine learning.

Cost ineffective-while using any electronic device there maybe a situation in which the electronic device may fail. The maintenance of the devices should be done from time to time.

Sensor working state-The improper working of sensor will result in lack of data transfer within the cloud

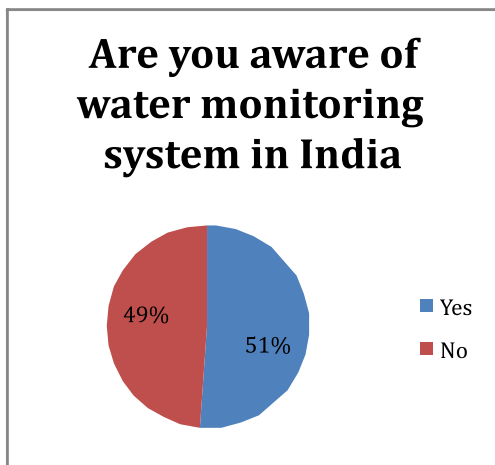
Example –Considering the sensor B in between sensor A and sensor C which is not working properly it will be difficult to know the area which is contaminated due to distance.

Varying the flow of river-During rainfall the level of water increases that results in the upstream and downstream of water. Since sensor works on limited flow of water, it will be difficult for sensor in detecting.

Findings and Observations:

Through this research study, the following observations were made by us:-

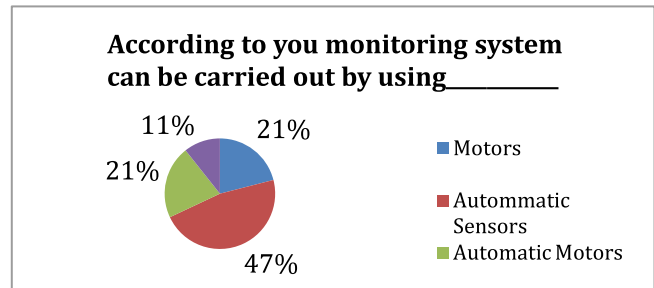
1. Are you aware of any water monitoring system in India? We observed that around 49% of people still don't know about the water monitoring system in India.



Options	No. of People
Yes	40
No	37

2. According to you, monitoring system can be carried out by using_____?

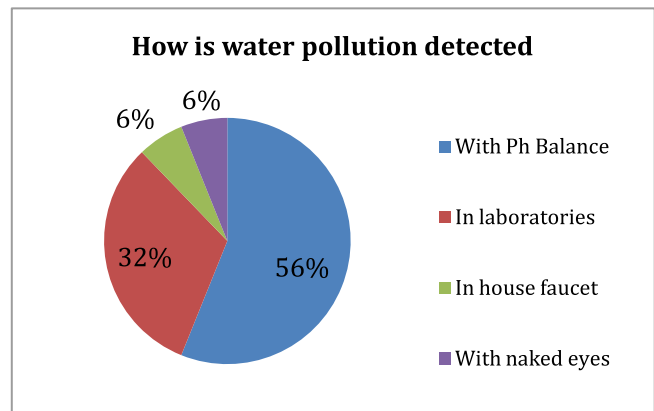
We observed that many people think that automatic sensors will be most useful tool to carry out water monitoring system.



Options	No. of People
Motors	6
Automatic Sensors	43
Automatic Motors	6
Turbines	18

3. How is water pollution detected?

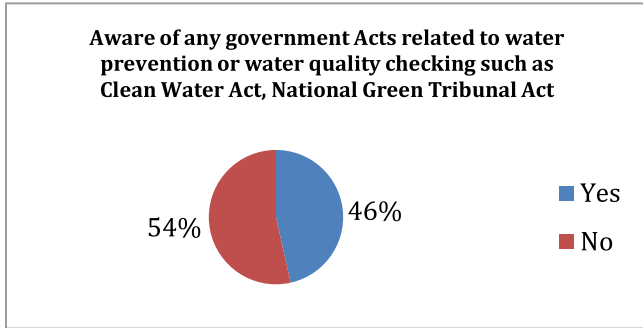
We observed that around 56% of people know the parameter required for detection of water pollution.



Options	No. of People
With pH balance	49
In laboratories	21
In house faucet	3
With naked eyes	2

4. Are you aware of any government Acts related to water prevention or water quality checking such as the Clean Water Act, or the National Green Tribunal Act?

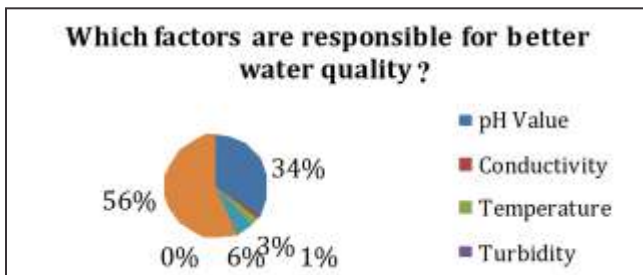
We observed that more than 56% of people are not aware with any of government Acts related to water quality checking and its prevention.



Options	No. of People
Yes	43
No	34

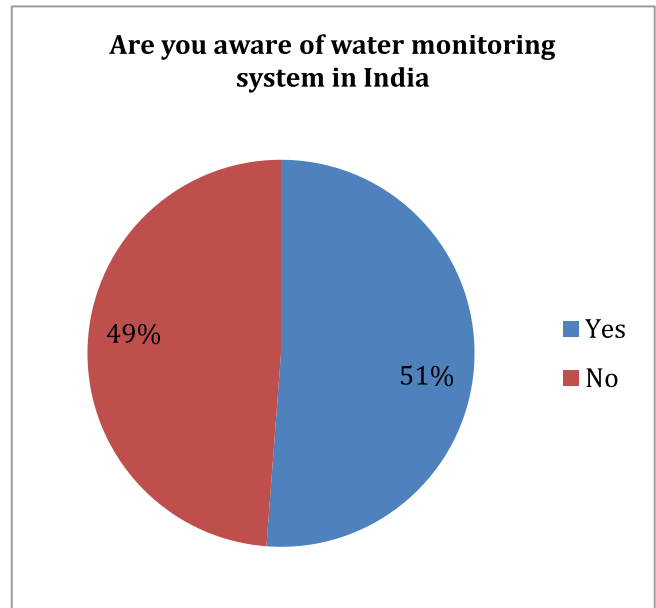
5. Which factors are responsible for better water quality?

We observed that 34% of people think that pH value of water is responsible for better water quality.



Options	No. of People
pH Value	25
Conductivity	1
Temperature	2
Turbidity	0
Dissolved Oxygen	4
All of the above	41

6. Are you aware of any water monitoring system in India? We observed that around 49% of people still don't know about the water monitoring system in India.



Options	No. of People
Yes	40
No	37

Future Scope :

Artificial intelligence, Internet of things (IOT), Machine learning all these technologies act as fundamental building blocks to make the tasks easier in the near future and can have direct impact on human health.

Reinforcement Machine learning is a type of machine learning which applied as water quality monitoring technology to enhance prior machine learning along with sensor in many ways. Various applications such as cyber security, data analysis, transport, self-driving cars are already being witnessed.

Also the Marketing advertising is benefitting the business as the best recommendation systems possible used.

Machine learning technology along with its further enhanced technology like deep learning which uses **Convolutional Neural Network** and re-enforcement learning can also be used in the near future to make the water monitoring even more easier and applicable to many different forms not even in detection and monitoring but also in various fields of technology.

Conclusion :

The proposed system of water monitoring using machine learning along with the sensor technology and the algorithm responsible to detect the anomalous behaviour using unsupervised machine learning done very well. On plotting the graph by using the dataset of Yamuna river in year 2013-2014, which is the most polluted river in India. Considering some water parameters as important factors to detect the anomalous behaviour of the water of 2 years, we found the variation trends responsible for water contamination due to the change in environment and the sensors applied at appropriate location under water able to get the contamination source approximately well.

On the basis of the survey, it is concluded that many people do not even know about the techniques/technology used for the water quality monitoring which have a major impact directly on human health. Different water Act implemented under Indian government do not make any difference in handling water quality purity.

Hence, Machine learning is the way through which technology is changing our lives for a better future and also having better scope to have a great impact in our lives.

References :

Article Based On “Machine Learning Approaches for anomaly detection of water quality on real-world dataset” by Doina Logofatu and Florine Leon faculty of Computer Science and Engineering Frankfurt university of Applied Sciences, University Of Lasi.

Overview of Clean Water AI- using AI to detect dangerous bacteria and harmful particles in water- Team clean water AI-(Peter Ma, Justin Shenk, Shin Ae Hong, Sarah Han).

Web Links :

<https://www.usgs.gov/nwis-national-water-information-system>.

<https://www.iisd.org/sites/default/files/publications/water-quality-monitoring-system-design.pdf>.

https://www.researchgate.net/publication/287647835_Smart_Water_Quality_Monitoring_System

<https://www.azosensors.com/article.aspx?ArticleID=1444>

<https://www.quora.com/What-is-parameter-tuning-in-machine-learning>

<https://theberkey.com/blogs/water-filter/9-things-that-causes-water-pollution-and-its-possible-solution>

<https://data.gov.in/resources/water-quality-river-yamuna-2013>

<https://data.gov.in/resources/water-quality-river-yamuna-2014>

<https://towardsdatascience.com/anomaly-detection-with-isolation-forest-visualization-23cd75c281e2>

<https://data.gov.in/resources/water-quality-river-yamuna-2012>

<https://www.kdnuggets.com/2018/03/5-things-reinforcement-learning.html>