**CHAPTER 1**

**INTRODUCTION**

Water is the most important resource for life, as it is required for the survival of living organisms and humans. From The United Nations world water development report 2017, approximately 80% of the world's wastewater is discharged back into the environment, largely untreated, damaging rivers, lakes, and seas. The widespread problem of water pollution is endangering our health. If nothing is done, the problems will only worsen by 2050, when global freshwater demand is estimated to be one-third more than it is currently.

Water quality monitoring through classification has become an important method to control the pollution level. A manual calculation is used to determine the water class when classifying water quality. The manual analysis is inefficient since it takes a long time to complete the operation. As a result, an improved system to classify water quality is necessary. An effort to ensure water quality is the development of a water quality monitoring system.

By utilizing parameters such as the pH, Hardness, Solids, Conductivity, Sulphates, Chloramines, Trihalomethanes, Organic Carbon, the system can calculate and classify the pollution of the water to a certain degree . The process involves using an algorithm to calculate the data collected from rivers or lakes.

Support vector machines (SVM) is a kernel approach used to solve classification and regression problems. Hyperplanes are used by SVMs, which were originally developed for binary classification, to construct decision boundaries between data points of different classes. The SVM algorithm is good at generalizing the sample, dealing with nonlinear problems in a simple but highly accurate way, and differentiating data that differs from the norm; therefore, it's ideal for classifying and evaluating water quality data.

Extreme gradient boosting or XGBoost is a gradient boosting framework from a decision-tree-based ensemble Machine Learning algorithm. XGBoost can be used to solve problems involving regression, classification, ranking, and user-defined prediction. In various machine learning and data mining challenges, the impact of XGBoost has been widely recognized, even for water classification.

Water quality needs to be studied comprehensively because of its importance in daily life and its effects on human health. Water quality monitoring offers the objective data needed to make informed decisions about water quality management now and in the future. As a result, water quality must be monitored to ensure that no contaminants exceed levels that are hazardous to human health. Water quality is currently determined through costly and time-consuming lab and statistical analyses. It necessitates sample collection, transportation to labs, and a significant amount of time and calculation, which is ineffective given that water is a highly communicable medium and hazardous if water contaminated with disease-causing waste is not prevented earlier . In this case, an alternative method based on machine learning for the efficient prediction and classification of water quality levels is implemented. We propose the implementation of SVM and XGBoost as an alternative method for water quality classification.

* 1. **Internet of Things**

The Internet of Things (IoT) refers to the interconnectivity of physical devices and objects through the internet, allowing them to collect and exchange data without human intervention. This network of connected devices includes everything from smartphones and smart homes to industrial machinery and city infrastructure. The IoT has the potential to transform numerous industries, including healthcare, transportation, agriculture, and manufacturing, by increasing efficiency, reducing costs, and enabling new services and products. However, the proliferation of IoT devices also raises concerns around privacy, security, and data management.

* + 1. **Purpose of Internet of Things**

The purpose of using the Internet of Things (IoT) in water quality prediction is to enable real-time monitoring, analysis, and management of water resources. IoT devices, such as sensors and smart meters, can be deployed in water bodies, treatment plants, and distribution networks to collect and transmit data on water quality parameters, such as pH, temperature, dissolved oxygen, and turbidity.

In addition to water quality prediction, IoT can also contribute to more efficient water management practices by enabling better resource allocation, reducing water waste, and improving the accuracy of billing and metering. Overall, the purpose of IoT in water quality prediction is to promote sustainable and effective water management practices that ensure the safety and reliability of our water resources.

* 1. **Machine Learning**

Machine learning is a subset of artificial intelligence that involves the use of algorithms and statistical models to enable computers to learn from data, without being explicitly programmed. Machine learning systems can identify patterns, make predictions, and take actions based on the data they receive.

There are three main types of machine learning: supervised learning, unsupervised learning, and reinforcement learning. Supervised learning involves training a model on a labeled dataset, where the output is known, to predict the output for new data. Unsupervised learning involves discovering patterns and relationships in unlabeled data, without specific target variables. Reinforcement learning involves training a model to take actions in an environment to maximize a reward.

Machine learning has numerous applications in various industries, including healthcare, finance, transportation, and marketing. It is used to detect fraud, diagnose diseases, and personalize user experiences.

However, the effectiveness of machine learning models depends on the quality and quantity of data used to train them, as well as the algorithms and techniques used. Additionally, ethical considerations such as bias and privacy must be taken into account when implementing machine learning systems.

**1.2.1 Purpose of Machine Learning**

The purpose of using machine learning in water quality prediction is to improve the accuracy and efficiency of predicting and monitoring water quality parameters. Machine learning algorithms can analyze vast amounts of data from sensors and other sources to identify patterns and trends that may be difficult for humans to detect.

* 1. **Description of the project**

This project aims to develop a water quality prediction system by leveraging the power of machine learning and internet of things (IoT) technologies. The system will consist of multiple sensor nodes that monitor various water quality parameters such as temperature, pH, solids, hardness, chloramines, trihalomethanes, organic carbon, sulphates, conductivity and turbidity.

Machine learning algorithms are then applied to the collected data to analyze it and make predictions about the water quality. The system can be trained to detect abnormal changes in the water quality and alert the authorities in case of potential water pollution or contamination.

The system can also be used to provide insights into the water quality and help authorities make informed decisions on water management and treatment. For instance, the system can predict the amount of chlorine required to treat the water, which can help in reducing the costs of water treatment.

Overall, the Water quality prediction system using IoT and Machine learning project can help in the management and protection of water resources by providing real-time data on water quality and predicting potential pollution events.

**1.4 Problem Analysis**

One of the main challenges in water quality management is the timely and accurate detection of contaminants. Traditional methods of water quality monitoring involve manual sampling and laboratory analysis, which can be time-consuming, expensive and prone to errors.

The current water quality monitoring systems typically provide only periodic measurements at discrete locations, which can result in missed events or insufficient spatial and temporal coverage.

The IoT sensors and connectivity technologies for real-time monitoring of water quality need to be deployed and configured properly to capture accurate and reliable data. With the large amounts of data generated by IoT sensors, it can be challenging to process, store, and analyze the data in real-time.

Machine learning algorithms can be used to predict water quality levels based on historical data, but they require large amounts of reliable and diverse data for accurate predictions.

The IoT sensors and the machine learning algorithms need to be maintained and updated regularly to ensure their accuracy and reliability over time.