IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 4, April 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204010 |

Water Quality Prediction using Machine Learning Algorithms

Devibala Subramanian[1], Rahmathnisha A[2]

[1]Assistant Professor, PG& Research Department of Computer Science, Sri Ramakrishna College of Arts & Science, Coimbatore, Tamil Nadu India

[2]UG Student, PG& Research Department of Computer Science, Sri Ramakrishna College of Arts & Science, Coimbatore, Tamil Nadu India

ABSTRACT: Water quality prediction is a critical aspect of environmental monitoring and management. Ensuring the availability of safe and clean water is essential for human health and the sustainability of ecosystems. In this study, we propose a machine learning-based approach to predict water quality parameters. The developed model leverages historical water quality data, meteorological information, and other relevant features to forecast key indicators such as pH, dissolved oxygen, turbidity, and chemical concentrations. The dataset used for training and validation is sourced from various water bodies, encompassing diverse environmental conditions. Feature engineering techniques are employed to extract meaningful information from the raw data, and machine learning algorithms, including regression and ensemble methods, are utilized to build predictive models. The model's performance is evaluated through cross-validation and compared against traditional statistical methods. The predictive system offers a valuable tool for early detection of potential water quality issues, enabling timely intervention and resource allocation. The integration of real-time data from sensor networks further enhances the model's adaptability to dynamic environmental changes. Our findings demonstrate the efficacy of machine learning in water quality prediction, contributing to the advancement of smart water management systems and the preservation of water resources. This research has implications for policymakers, environmental scientists, and water utility managers seeking efficient and proactive strategies for maintaining water quality. The data is used monthly and daily on our mobile phone. The mobile data package is used day by day to recharge the data pack. A data saver is also used for mobile data. The data is used for one day, but it is not full data. It is used as mobile data in a mobile data package. They will continue in the next day's data package, and the data will be saved and stored in our mobile system storage. The data-saving mode will be on your system. The data saved by the data package network in your mobile data stored, a network used the next day will continue. If your data package of 1 GB is used in one day, you will have 500 MB left. On balance, 500 MB are used for the next-day usage data package on mobile data. Used by data in one day and balance data stored an automatic mobile stored in our mobile network data, and that data is used by the next day also using our mobile network. If you used 600 MB and the balance of 400 MB is stored automatically in our mobile, the next day you will continue to use a data package. If, suppose, in one day the data package is not used in full and a half day is used, the balance data package is next and will continue in our data package.

KEYWORDS: Data saving mobile network, Data analysis techniques, Wi-Fi data background

I. INTRODUCTION

Water quality prediction is crucial for environmental monitoring and management, ensuring safe water for human health and ecosystem sustainability. This study proposes a machine learning-based approach to predict water quality parameters, utilizing historical data, meteorological information, and relevant features to forecast indicators like pH, dissolved oxygen, turbidity, and chemical concentrations. The dataset, sourced from diverse water bodies, undergoes feature engineering to extract meaningful information, and machine learning algorithms such as regression and ensemble methods construct predictive models. Cross-validation evaluates model performance against traditional methods, offering an early detection tool for water quality issues and facilitating timely interventions. Integration with real-time sensor data enhances adaptability to environmental changes. Our findings highlight machine learning's efficacy in water quality prediction, benefiting smart water management and resource preservation. Policymakers, environmental scientists, and water utility managers can leverage these insights for proactive strategies in maintaining water quality.

Water quality prediction through machine learning involves using algorithms to assess various parameters and

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204010 |

anticipate water quality levels. Initially, historical data on water quality is gathered from diverse sources like sensors, laboratories, and online repositories. Subsequently, pertinent factors such as pH levels, turbidity, dissolved oxygen, temperature, and pollutant concentrations are chosen for modelling. These components function as input variables for the machine learning model. Commonly used supervised learning algorithms for prediction tasks include regression, decision trees, random forests, or neural networks. The dataset is partitioned into training and testing sets to evaluate the model's accuracy effectively. Feature engineering methods may be applied to amplify predictive capabilities. The model is then trained on the training data to understand the patterns and connections between the input features and water quality parameters. Hyperparameter tuning is carried out to enhance the model's performance. Cross-validation methods are used to determine the model's ability to generalize. Once the model is trained, it is validated using the testing dataset to ensure its capacity to predict water quality accurately in unseen data. Performance metrics such as accuracy, precision, recall, and F1-score are computed to gauge the model's efficiency. Continuous monitoring and refinement of the model are crucial to adapt to shifting environmental conditions. Integration with Internet of Things (IoT) devices enables real-time monitoring and prognostication of water quality. Predictive models contribute to the early identification of water contamination or irregularities, facilitating prompt interventions to ensure water safety. Ultimately, machine learning-based water quality prediction bolsters environmental oversight and public health by furnishing pre-emptive measures for sustaining and enhancing water quality

II. RELATED WORKS

Seema[1] proposed water resources are equally important for natural ecosystem and human development. It is essential for agriculture, industry and human existence. All life on earth depends on water. Fresh water is a critical, finite, vulnerable, renewable natural resource on the earth and plays as important role in our living environment without it life is impossible. More than 70% of the Earth's surface is covered with this simple molecule. Scientists estimate that the hydrosphere contains about 1.36 billion cubic kilometers of these substances of physicochemical parameters such as pH, color, biochemical oxygen demand (BOD), chemical oxygen demand (COD), total suspended solids (TSS), total dissolved solids (TDS), and turbidity. The quality of water can be assessed by studying its physical and chemical characteristics.

Gordge[2] proposed water is the most important in shaping the land and regulating the climate. It is one of the most important compounds that profoundly influence life. The quality of water usually described according to its physical, chemical and biological characteristics. Rapid industrialization and indiscriminate use of chemical fertilizers and pesticides inagriculture are causing heavy and varied pollution in aquatic environment leading to deterioration of water quality and depletion of aquatic biota. Due to use of contaminated water, human population suffers from water borne diseases.

Pawari and Gavande[3] proposed Water is one of the vital needs of all living beings. Humans need water in many daily activities like drinking, washing, bathing, cooking etc. If the quality of water is not good then it becomes unfit for drinking and other activities. The quality of water usually described according to its physical, chemical and biological characteristics. Hence it becomes necessary to find the suitability of water for drinking, irrigation and Industry purpose. The groundwater quality based on Sodium percent, Sodium Absorption Ratio and Residual Sodium Carbonate will help to identify the suitability of water for irrigation purpose. Rapid industrialization and use of chemical fertilizers and pesticides in agriculture are causing deterioration of water quality and depletion of aquatic biota. Due to use of contaminated water, human population suffers from water borne diseases. Parameters that may be tested include temperature, pH, turbidity, salinity, nitrates, TDS, Cations, Anions and phosphates. Total Hardness was observed some evidence indicates its role in heart diseases and hardness of 150-300 mg/l and above may cause kidney problems and kidney stone formation, as it causes unpleasant taste and reduce ability of soap to produce lather. Hard water is unsuitable for domestic use. The suggested measures to improve the ground water quality includes total ban on the activities that causes pollution, avoid use of pesticides and prevent the sea wage.

Jamie and Richard[4] proposed the method for determining the total phosphorus concentration in water sample as it is described by Jamie Bartram and Richard Balance () consists of converting all organically combined phosphorus and all phosphates to reactive by digestion in Presence of potassium proxy di sulphate . Then the reactive phosphorus is analyzed by colorimetric method as described for reactive phosphate.

George and Edward[5] proposed Surface waters are subject to temperature variation due to fluctuation in sunlight and air temperature. Temperature was slightly higher in the afternoon compared to the morning due to longer exposure to sunlight. The temperature difference between surface and bottom can be explained by the difference in light penetration. Temperature recorded at different stations did not show big variation during the same sampling period which indicates uniform conditions with respect to temperature.

Sarala and Ravi[6] proposed the groundwater quality based on Sodium percent, Sodium Absorption Ratio and Residual Sodium Carbonate will help to identify the suitability of water for irrigation purpose. Rapid industrialization

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204010 |

and use of chemical fertilizers and pesticides in agriculture are causing deterioration of water quality and depletion of aquatic biota. Due to use of contaminated water, human population suffers from water borne diseases.it can been generated in the levels of functional review of the water quality and it can be used many parameters in this section.

Heidtke and Asce[7] proposed water is the necessity of any living species on the earth. We even cannot imagine our life without unavailability of water in our daily routine. Our body contains 70% of water inside it and it plays a crucial role to make our body healthy and workable. But the public places like railway stations, bus stations, public parks, etc. the water facility is not proper and not fit for drinking purpose. A study was carried out to evaluate the quality of drinking water supplied at major railway stations in south Gujarat. Water sample were connected from the tap water sources making at total of 2 sampling points of south Gujarat railway station. There were Bharuch , surat , valsad and vyara railway stations.

Pawari and Sagar[8] proposed water is one of the most common aspects to evaluate the quality we are using certain parameters . Present work is analysis of ground water pollution in hadapsar region. This area has got many open wells which are connected by ground water table of surrounding area. As the water is used for drinking and irrigation purposes, it becomes essential to find the suitability of these open wells for drinking and irrigation purpose. In current work sample from 9 Open wells from study area were Selected and tested for various parameters. The hydro chemical analysis has been done by using piper diagram for both the seasons, i.e. pre and post monsoon. The groundwater quality is tested based on Sodium percent, Sodium Absorption Ratio and Residual Sodium Carbonate & suitability of water for irrigation purpose is examined and found out it is suitable for irrigation purpose. From the result it has shown the sampling points having hardness, TDS, conductivity and chlorides exceed the permissible limit as per APHA standards .So that it is not safe for drinking purpose. Thus from the overall analysis some suggestions & remedial measures are provided in the paper for the same.

Bruce[9] proposed a systems, holistic, or ecosystem approach is often advocated for water management, and has led to the emergence of integrated water resource management, or IWRM. Such an approach can be interpreted as `comprehensive' or `integrated', and analysts, planners, and managers need to under-stand the difference.Edge or boundary problems always are encountered when applying a holistic approach, and design of institutional arrangements cannot eliminate these problems but can minimize them. IWRM often does not have a statutory basis, which can lead to implementation challenges.By linking IWRM to land-use planning and official plans at the local level, IWRM can be given credibility, as well as be systematically connected to land-based issues.

Hussaina and maneb[10] proposed the present work we are reported the Physicochemical properties like pH, conductivity, Turbidity, TDS, DO, fluoride, chloride, Sodium, Sulphate , etc. and the values are compared for treated and untreated water samples. The samples were collected from treatment plant of Ahmadpur, Dist Latur.t he values changes apparently after the treatment of water.

Pradhan and mohsin[11] proposed that Water sample were connected from the tap water sources making at total of 2 sampling points of south Gujarat railway station. There were Bharuch , surat , valsad and vyara railway stations.Two set of sample collected,which is one during monsoon and other one is after the monsoon where taken from each sampling point.then it can can be value the substances and performance of the analyzing parameter in chilika lake water.

Devendra and shriram[12] proposed that human and industrial activities the ground water is contaminated. This is the serious problem now a day. Thus the analysis of the water quality is very important to preserve and prefect the natural eco system. The assessment of the ground water quality was carried out in the different wards of Indore City. The present work is aimed at assessing the water quality index (WQI) for the ground water of Indore City andits industrial area .The ground water samples of all the selected stations from the wards werecollected for a physiochemical analysis. For calculating present water quality status by statistical evaluation and water quality index, following 27 parameters have been considered Viz. pH, color, totaldissolved solids. electrical conductivity, total alkalinity, total hardness, calcium, chromium, zinc, manganese, nickel.The obtained results are compared with Indian Standard Drinking Water specification IS: 10500-2012. The study of physico-chemical and biological characteristics of this ground water sample suggests that the evaluation of water quality parameters as well as water quality management practices should be carried out periodically to protect the water resources.

Mohammed and [13] remotely sensed data can reinforce the abilities of water resources researchers and decision makers to monitor water bodies more effectively. Remote sensing techniques have been widely used to measure the qualitative parameters of water bodies (i.e., suspended sediments, colored dissolved organic matter (CDOM), chlorophyll-a, and pollutants). A large number of different sensors onboard various satellites and other platforms, such as airplanes, are currently used to measure the amount of radiation at different wavelengths reflected from the water's surface. In this review paper, various properties (spectral, spatial and temporal, etc.) of the more commonly employed space borne and airborne sensors are tabulated to be used as a sensor selection guide. Furthermore, this paper investigates the commonly used approaches and sensors employed in evaluating and quantifying the eleven water quality parameters.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204010 |

III. PROPOSED WORK

Water quality prediction is an essential aspect of ensuring public health and safety. Random Forest algorithm is a powerful tool that can be used to analyze water quality data to predict future water quality. By leveraging the power of machine learning, this system can provide accurate and reliable predictions of water quality, which can help ensure public health and safety. In this project we have compared four machine learning algorithms Such as Random Forest, Naive bayes, Decision Tree and Logistic Regression Algorithm. A Random Forest Algorithm is a supervised machine learning algorithm that is extremely popular and is used for Classification and Regression problems in Machine Learning. We know that a forest comprises numerous trees, and the more trees more it will be robust. Similarly, the greater the number of trees in a Random Forest Algorithm, the higher its accuracy and problem-solving ability. Random Forest is a classifier that contains several decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset. It is based on the concept of ensemble learning which is a process of combining multiple classifiers to solve a complex problem and improve the performance of the model.

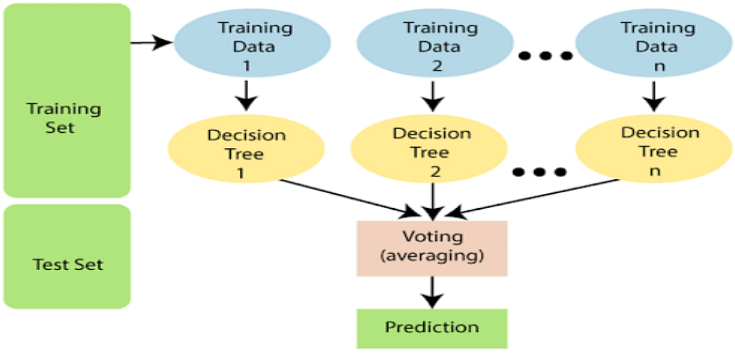

Figure-1 Working of Random Forest Algorithm

The following steps explain the working Random Forest Algorithm:

Step 1: Select random samples from a given data or training set.
Step 2: This algorithm will construct a decision tree for every training data.
Step 3: Voting will take place by averaging the decision tree.
Step 4: Finally, select the most voted prediction result as the final prediction result.

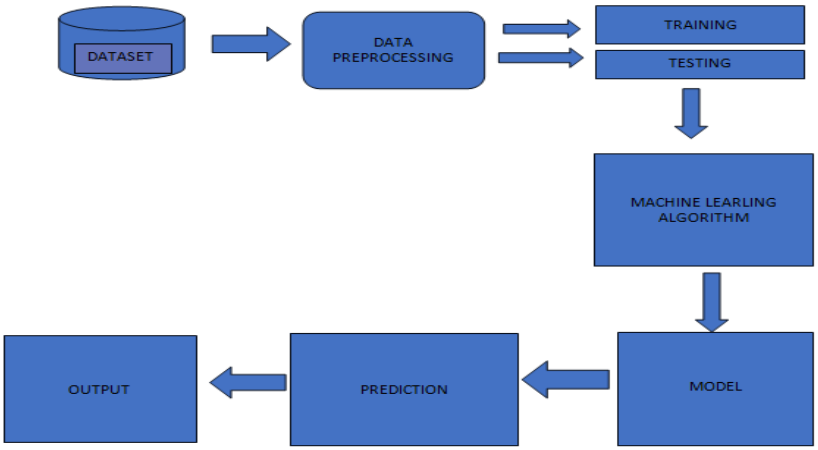

Figure-2 System Architecture

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204010 |

IV. RESULTS AND DISCUSSION

In the results section, you would typically present the performance metrics of your machine learning models, such as accuracy, precision, recall, and F1-score. You may also include visualizations, such as confusion matrices or ROC curves, to illustrate the model's performance. Additionally, you could discuss any insights gained from analysing the data and model predictions.

1. Model Performance: Analyse the performance of different machine learning models and discuss which models perform best for predicting water quality.
2. Feature Importance: Discuss the importance of different features in predicting water potability and any insights gained from feature analysis.
3. Limitations: Address any limitations of your study, such as data quality issues, imbalanced classes, or simplifying assumptions made during model training.
4. Generalization: Consider the generalizability of your model to unseen data and discuss potential challenges in deploying the model in real-world scenarios.
5. Ethical Considerations: Reflect on any ethical considerations related to predicting water quality and how the model's predictions might impact different stakeholders.
6. Future Directions: Propose potential avenues for future research or improvements to the model, data collection process, or deployment strategy.

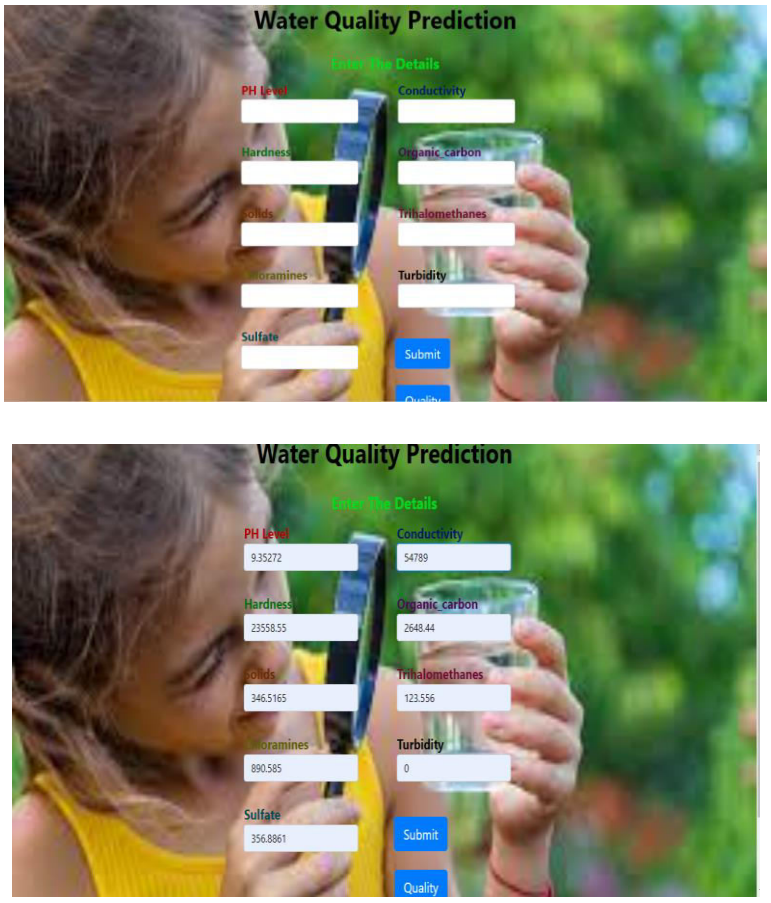

Figure 3: Hydraulic Purity Forecasting Test Form

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204010 |

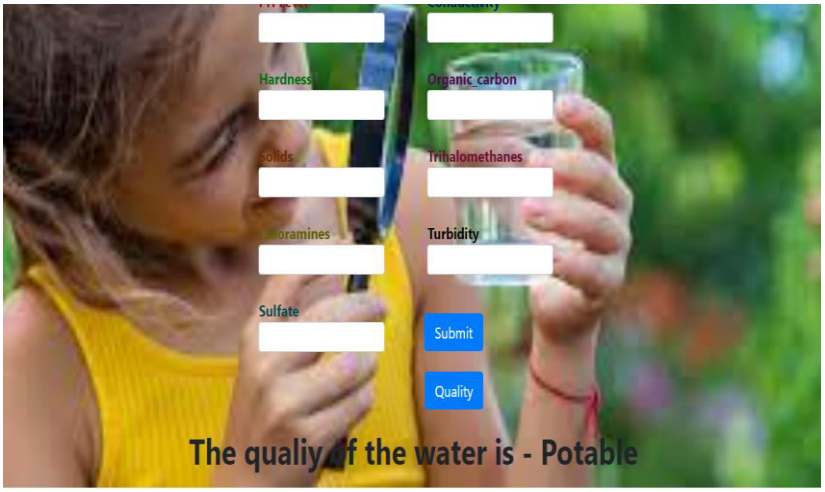

Figure-4 Classification Results

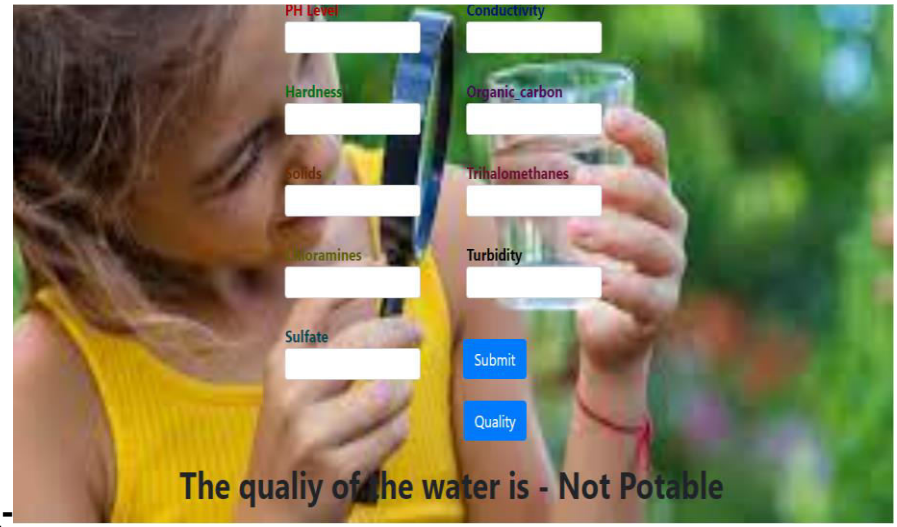

Figure-5 Classification Results

V. CONCLUSION

In conclusion, water quality prediction is an important problem in environmental science, with applications in water resource management, public health, and ecosystem conservation. Machine learning algorithms have shown promise in accurately predicting water quality based on various physicochemical and biological parameters. Different machine learning algorithms such as random forest, logistic regression, decision tree, and naive Bayes have been applied successfully in water quality prediction. However, the choice of algorithm depends on the characteristics of the dataset and the specific prediction problem. To achieve accurate and reliable water quality prediction, it is important to collect and analyze high-quality data, preprocess the data to handle missing values and outliers, select appropriate features, and evaluate the performance of the models using appropriate metrics.

REFERENCES

[1] Dr.Seema Tiwari, “Water Quality Parameters”, Sagar Iinstitute of Research and Technology, Bhopal, India, International Journal of Engineering Science Invention Research & Development

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204010 |

[2] S.P.Gordge,M.V.Jadhav, "Assessment of Water Quality Parameters", S.P.C.O.E.M.R.,College of Engineering, Pune, University of Pune, Int. Journal of Engineering Research and Applications,ISSN : 2248-9622, Vol. 3, Issue 6, Nov-Dec 2013, pp.2029-2035.

[3] M.J.Pawari1, S.M.Gavande2, "Assessment of Water Quality Parameters", P.G.Scholar, Department of Civil Engineering, Anantrao pawar College of Engineering and Research, Parvati, Pune, Maharashtra, India , International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064 Index Copernicus Value (2013): 6.14 | Impact Factor (2013): 4.438 Volume 4 Issue 7, July 2015.

[4] Jamie Bartram and Richard Balance. (1996),"Water Quality Monitoring",United Nation Environmental Program & World Health Organization, 376p © 1996 UNEP/WHO ISBN 0 419 22320 7 (Hbk) 0 419 21730 4 (Pbk).

[5] George Tchobanoglous & Edward D Schroeder. (1987). "Water quality: characterization,modeling, modification",706p,University of California.

[6] Sarala C, Ravi Babu P, " Assessment of Groundwater Quality Parameters in and around Jawaharnaga", Hyderabad, International Journal of Scientific and Research Publications, Volume 2, Issue 10, 2012, pp.1-6 [7] T. M. Heidtke, A. M. Asce and W. C T. M. Heidtke, A. M. Asce and W. C. Sonzogni ,"Water Quality Management for the Great Lakes", J. Water Resour. Plann. Manage. 112:48-63, 1986.

[8] M.J.Pawari& Prof. Sagar Gawande, "Assessment of Underground water Quality around Hadapsar region in Pune", Maharashtra., International Research Journal of Engineering and Technology (IRJET), Volume: 02Issue: 04| July-2015.

[9] Bruce Mitchell, "Water quality analysis and parameters", Department of Geography, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada.

[10] S.Hussaina, V.Maneb, STakdea, A. Pathanc, M. Farooquic,"Comparison between Treated and Untreated water so as to study water treatment plant of Ahmadpur Dist. Latur," International Journal of Modern Engineering Research (IJMER) www.ijmer.com, Vol.1, pp-564-569, ISSN: 2249-6645

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details