Detection Of Water Quality Using Machine Learning

[1]Suma K, [2]Akash K S, [3]Dheeraj D R, [4]Jeevan Kumar, [5]Karthik N P

[1]Assistant Professor, [2,3,4,5]Student

[1,2,3,4,5]Computer Science and Engineering

[1,2,3,4,5]Mangalore Institute of Technology and Engineering, Moodabidri, Karnataka, India

Abstract: With an increasing population in India, usage of water also increases. Water plays an important role in almost every living organism's life. The water that is used for drinking must be of good quality. If the quality of the water is not good, it can cause many diseases. The pollution of water can be caused by multiple reasons. The aim of this project is to integrate machine learning models with IoT devices to measure the potability of water. The parameters of water that are used to measure the potability of water are pH, hardness of water, solids, chloramines, sulphate, water conductivity, organic carbon, temperature level, and turbidity. For each parameter, there is a particular range of values. For water to be potable, all of these parameters should be within the specified range. In this project, a machine learning model called Extremely Randomised Tree Classifiers (Extra Trees Classifier) is trained to predict the potability of the water. IoT devices are used to measure the required parameters in water, and then measured values are given to the Extra Trees Classifier to predict the portability of water.

Index Terms: Water Quality, Internet of Things (IoT), Machine Learning, Extremely Randomized Decision Trees, Potability of water.

I. INTRODUCTION

Over the decades, the primary and the most important concern has been the protection of water and its quality. Water makes up three-fourth of the surface of earth. The ocean contains 97 percent of the world's water and only 3 percent is freshwater. About 3 percent fresh water, 2.5 percent is not available and only 0.5 percent is available. Consumption of contaminated water can cause diseases, which can also lead to death. Human plays an important role in water pollution. Water is the most crucial thing for each and every living organism. The Quality of water is been determined by set of physical as well as chemical parameters. For a quality of water to be good, there is particular value range for each parameter which may be physical or chemical. If all parameter values are within the particular range, then the water quality is good. If the quality of water is not up to the mark, it should be treated using standard methods and then make use of it for drinking purposes. The traditional way of determining quality of water samples in laboratory takes lot of time and effort. Nowadays computer technologies have overshadowed the traditional techniques.

II. LITERATURE SURVEY

[1] Sanaa Kaddoura,(2021).Evaluation of Machine Learning Algorithm on Drinking Water Quality for Better Sustainability. ZU Scholars. This study explains the how variety of Machine Learning(ML) algorithm is used to predict the water quality. After that results of ML algorithms are compared and model with best result is used to predict the water quality.

[2] Zhao Fu,(2020). Water Quality Prediction Based on Machine Learning Techniques.UNLV Thesis.
As a result of this study classifies the water prediction problem into 4 categories based on the size of given dataset. Multiple ML models based on the model ANFIS are developed to increase the predicting performances for water prediction problem.First,to elimate errors of the model ANFIS in the test phase, stratified sampling technique is employed to reduce the irregular distribution of dataset.

[3] Theyazn H H Aldhyani, Mohammed Al-Yaari, Hasan Alkahtani, Mashael Maashi,(2020).Efficient Prediction of Water Quality Index (WQI) Using Machine Learning (ML) Algorithms.Hindawi.
This study discusses on using artificial intelligence (AI) algorithm to develop a model that can measure the quality of the water. the advanced AI algorithm can be used to predict the WQI. ML algorithm is used in classification of WQI data.

III. METHODOLOGY

This project is intended to determine the potability of water through the proposed system. The various parameters that are used in this project to detect the water potability are pH, Hardness, Solids, Chloramines, Sulfate, Conductivity, Organic Carbon, Temperature and Turbidity. IoT devices are used to measure some properties of water. The collected IoT data are compared with the trained Machine Learning model to predict the quality of the water. It will detect weather water is suitable for drinking purpose or it contains the impurities that is harmful for drinking. The methodology of this proposed system is divided into several steps. The steps are Dataset collection, Dataset preparation, Model training, Model optimization, Model evaluation, Water potability Prediction.
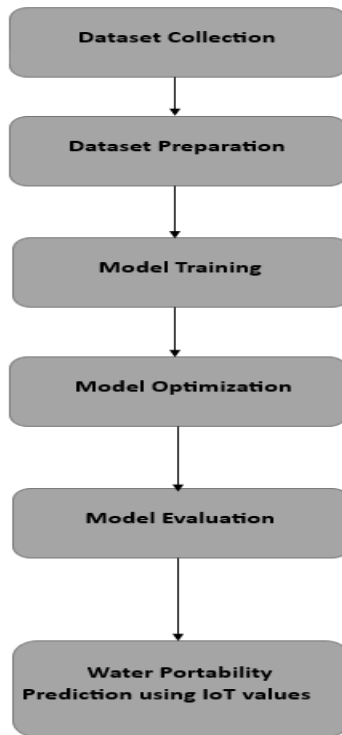
Figure 1: Flowchart of Methodology

A.*Dataset collection*

The Dataset is collected from multiple sources, that include Data world, Kaggle and etc.

B.*Data preparation*

Dataset collected from multiple sources are combined into a single dataset. The combined dataset is processed. The processing of the combined dataset includes replacing missing values in particular row-column with overall mean of that column.

C.*Model training*

This step includes training various machine learning models. The models which are trained are Random Forest, Decision Tree, SVM, Logistic Regression, Naïve Bayes, Extremely Randomized Trees Classifier. Based on Accuracy we have chosen Extremely Randomized trees classifier to predict the potability of water.

D.*Model Optimization*

This step includes optimization of the selected model. The Hyper parameter tuning using GridSearchCV is used for optimizing the model so that accuracy can increase. In Hyper parameter tuning we explicitly choose model parameters and specify values for that model parameters. The default set of model parameters is replaced with the new selected set.

E.*Model Evaluation*

In this step, we evaluate the selected model to check whether the accuracy of this model has increased.

F.*Water potability Prediction*

IoT devices are used to measure the physical parameters of water and chemical parameters are provided explicitly. If any parameter value is not known, it is replaced with mean value. The Extremely randomized trees Classifier takes as input the value given for parameters and predict the potability of water.
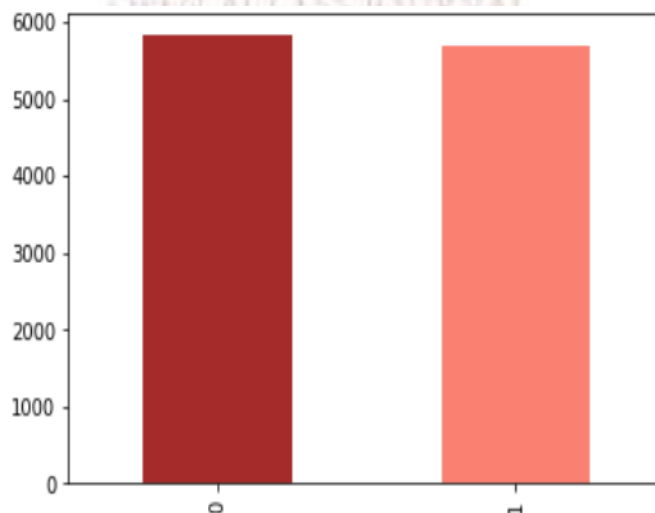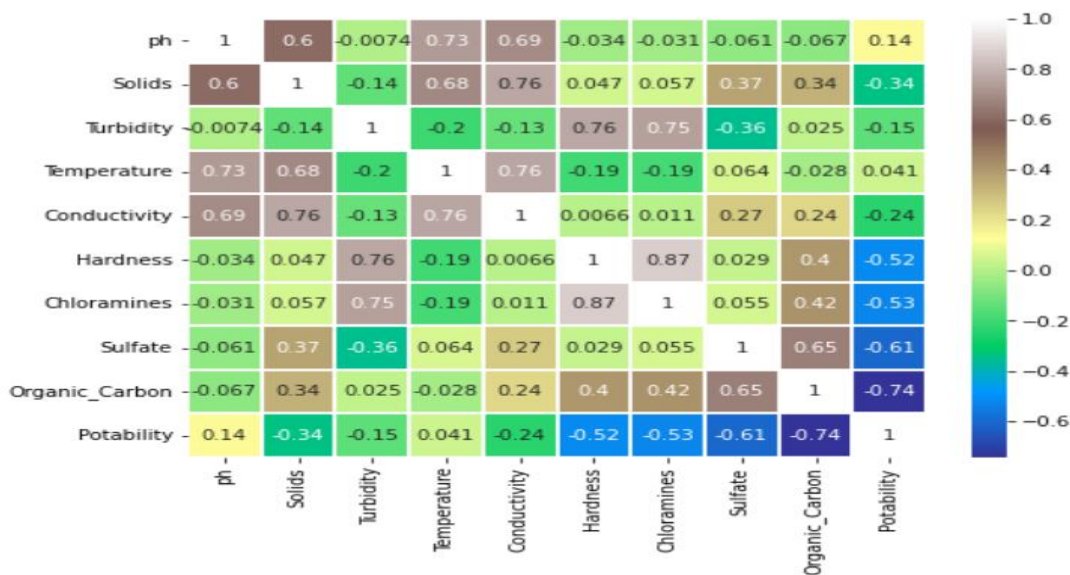

Figure 2: Potability Counts in Dataset

Figure 3: Correlation Heatmap

IV.RESULTS

Model Training phase involves training various models using the dataset prepared. The Models which are trained are Random Forest, SVM, Decision Tree, Naïve Bayes, Logistic Regression, Extremely Randomized Trees Classifier. The below Figure shows the Accuracy Comparison Between various models trained.
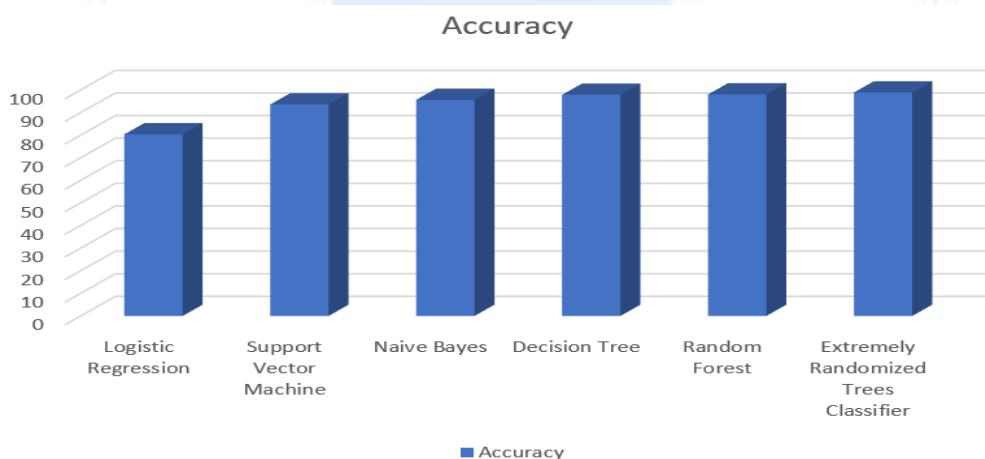

Figure 4:Accuracy Comparison Between Various Models

Based On the Accuracy,Extremely Randomized Trees Classifier is used to predict the water potability.Optimizing the selected model to increase the accuracy.After Optmizing,the Accuracy of the Extremely Randomized Trees Classifier has been increased from 98.91 to 99.
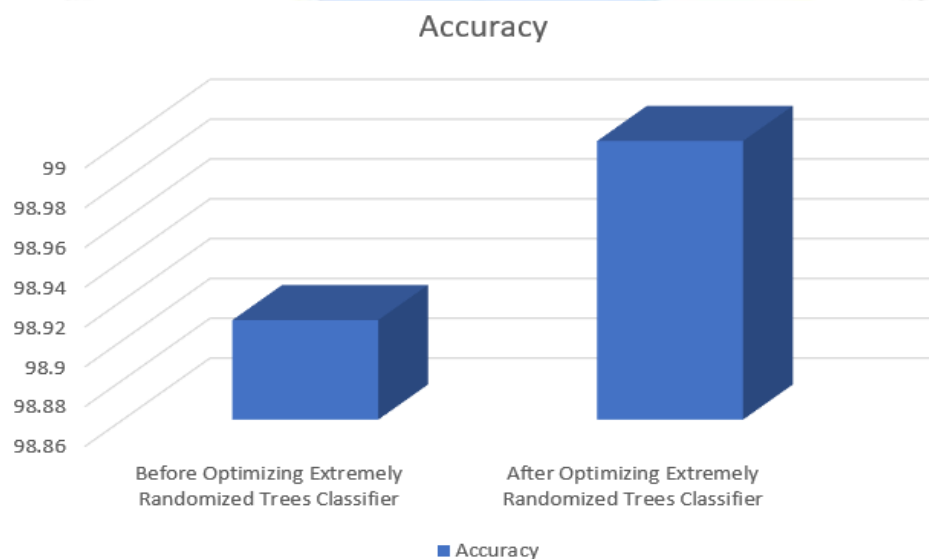

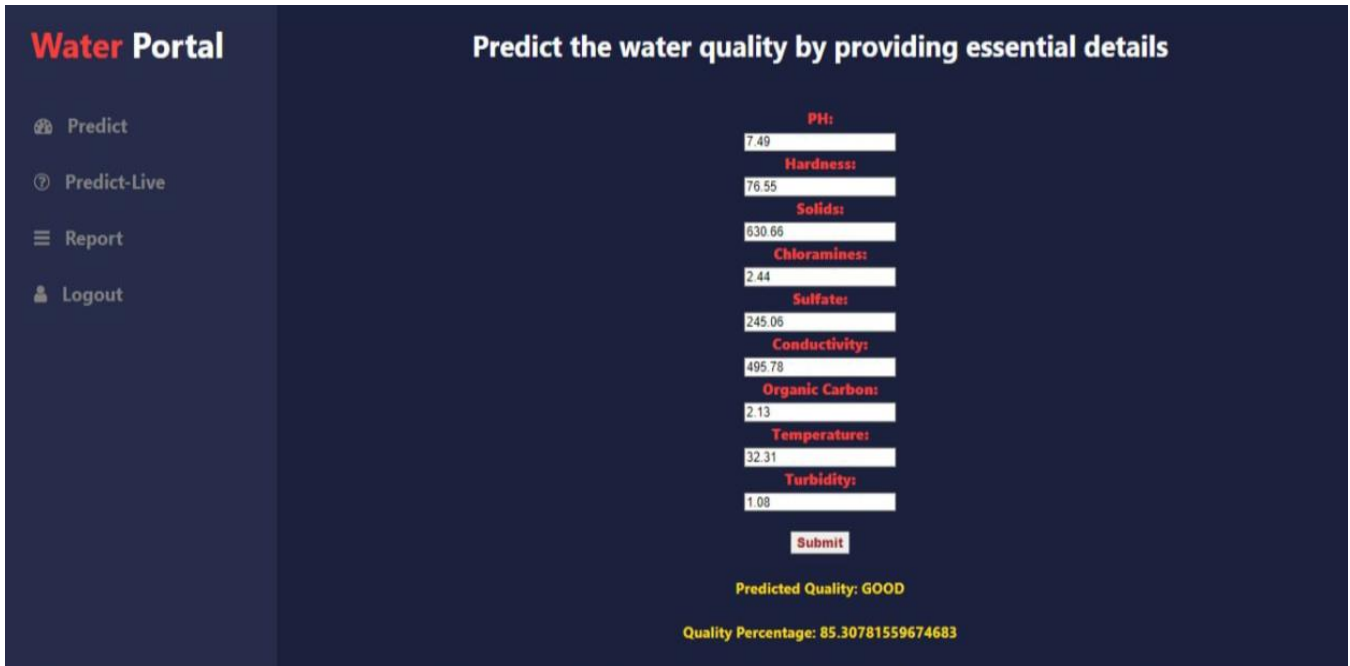Figure 5:Accuracy Comparison of Before and After Optimizing Extremely Randomized Trees Classifier
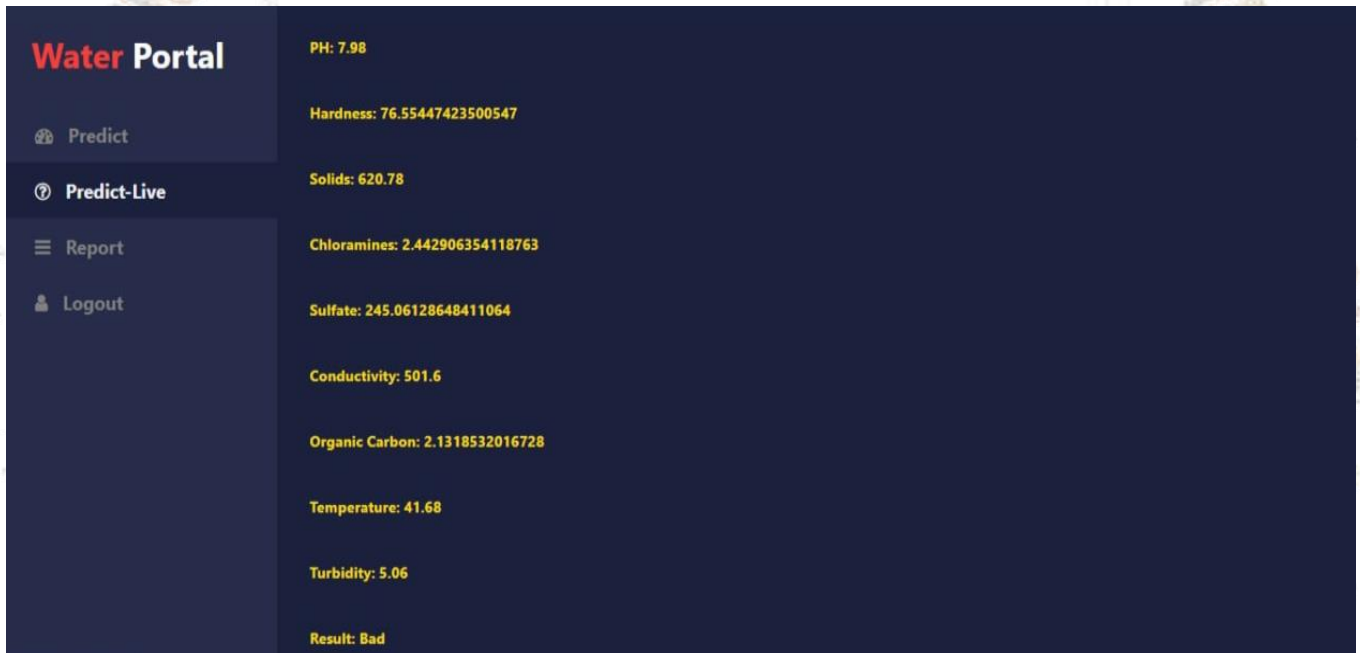
Figure 6: Predicting by filling the attribute values

Figure 7: Predicting using the IoT components

V.CONCLUSION

Water is the most important thing for every living organism. The quality of the water plays an important role, especially when it is used for drinking purposes. Water potability determines the water's quality. If water quality is good, it is safe to consume it. Testing the water quality in labs consumes more amount of time and efforts. In this project, a machine learning model is used to predict the water quality. This system detects the quality of the water and informs the appropriate users whether it is good or bad. Hopefully, this system will reduce the number of people who consume bad-quality water and reduce the risk of diseases.

VI.FUTURE SCOPE

In this project, IoT devices are used to measure only physical parameters of water. In future, IoT devices can also be used to measure chemical properties of water. The accuracy of the model used in this project can also be improved in the future.

VII.REFERENCES

[1] Sai Sreeja Kurra, Sambangi Geethika Naidu, Sravani Chowdala, Sree Chithra Yellanki, Dr. B. Esther Sunanda," Water Quality Prediction Using Machine Learning", Published in International Research Journal of Modernization in Engineering Technology and Science, Volume:04/Issue:05/May-2022, Impact Factor- 6.752, e-ISSN: 2582-5208.

[2] Manya Kakkar, Vansh Gupta , Jai Garg, Dr. Surender Dhiman," Detection of Water Quality using Machine Learning and IoT",Published in International Journal of Engineering Research & Technology (IJERT), ISSN: 2278-0181, Vol. 10 Issue 11, November-2021.

[3] Soundarya Pappu, Prathyusha Vudatha, Niharika.A.V, Karthick.T, Suresh Sankaranarayanan," Intelligent IoT Based Water Quality Monitoring System",Published in International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 16 (2017) pp. 5447-5454.

[4] Nida Nasir, Feras Barneih, Mustafa Sameer, Abdallah Shanableh, Ahmed Al-Shamma Water quality classification using machine learning algorithms

[5] U Ahmed, R Mumtaz, H Anwar, AA Shah, R Irfan, and J García-Nieto, Efficient water quality prediction using supervised machine learning, Water, Vol. 11, 2019, pp. 2210.