Water Purity Tester using Machine Learning Algorithms

Dr.B.Buvaneswari,
Associate professor,
Electronics and Communication Engineering,
K.L.N.College of Engineering, Sivagangai, India

Dr.T.R.Muthu,
Associate professor,
Electronics and Communication Engineering,
K.L.N.College of Engineering, Sivagangai, India

B.K.Priyadharshini,
UG Scholars,
Electronics and Communication Engineering,
K.L.N.College of Engineering, Sivagangai, India

T.K.Veera Sarika,
UG Scholars,
Electronics and Communication Engineering,
K.L.N.College of Engineering, Sivagangai, India

M.Rivenna Merlin
UG Scholars,
Electronics and Communication Engineering,
K.L.N.College of Engineering, Sivagangai, India

ABSTRACT-Water is a fundamental requirement for human, animal, and plant survival. Despite its importance, water is not always fit for drinking, domestic and industrial use. If drinking water contains unsafe levels of contaminants, it can cause health effects, which overall affects the social life and development. The main objective is to build a system that measures the quality of water. In this paper Embedded is integrated with Machine Learning for Data Accumulation and Assessment. PCA(Principle Component Analysis) is used to reduce the dimensionality of the predefined dataset. Several Algorithms namely K-nearest neighbor(KNN),Support Vector Machine(SVM),Decision Tree are compared and noted that Decision Tree has high accuracy. Machine Learning model has been trained using Existing dataset as Training data, The attributes such as pH, turbidity, Conductivity and total dissolved solid (TDS) obtained from the corresponding sensors are captured and turned into CSV file. The Real-Time data Analysis will be given by the Trained Machine Learning model.

KEYWORDS: Water Purity, Principle component Analysis(PCA), Decision Tree, NodeMCUESP8266,Turbidity Sensor ,PH Sensor, TDS sensor, Conductivity sensor.

I. INTRODUCTION

In the present days and current period, we are moving towards making our urban communities as the brilliant urban areas, because many innovative research and developments throughout the decades. So the present period is said to be time of creations, time of improvement, time of globalization and the time of astuteness and so on. In any case, the counter side of the equivalent is that the present time is time of the contamination, a dangerous atmospheric deviation, weakness and hopeless wellbeing factors. One of the inborn and prime hindrance is total population does not have purified and safe water for drinking. This is increasing unsafe circumstances in some nations like India, where grimy water is being utilized for drinking with no appropriate water treatment before drinking. The fundamental driver for this are the numbness of individuals

and government area and the inadequate water quality checking framework, which results in genuine medical problems.

The inspiration of the proposed framework was to plan a remote framework to screen water quality in a most straightforward and practical way. This framework can break down some essential variables of water to take preventive measures for water quality support. The pH sensor,turbidity sensor,conductivity sensor and total dissolved solid sensor (TDS) are utilized to gather the pH and turbidity dimension of the water. With the utilization of wifi module, we can get the information from the rural and less developed areas. The sensors have the simple yield, consequently they are interfaced to simple contribution of the Nodemcu microcontroller and the information are exchanged through the wifi module. The PH and turbidity parameters thus calculated are stored in a tabular format and is shown on pc. The parameters that are utilized to decide the nature of the water are the pH level and turbidity level.

The rest of the paper is organized as follows. The related works are explained in Section II. Section III, Section IV and Section V describes about the system architecture, software and hardware specification respectively. The existing system,proposed system and algorithms are discussed in Section VI and VII respectively.The result is discussed at section VIII. At last, Section IX concludes the paper with conclusion of the work.

II. RELATED WORKS

Said M.F. et al describes the possibility of submerged remote sensor organize is the water quality monitoring utilizing remote sensor arrange innovation controlled by sunlight based board [1]. The hubs and the base stations are associated utilizing WSN innovation like Zigbee Data assembled by various sensors at the center point side, for instance, pH, Information gathered from the remote site can be shown in visual configuration a well as it very well may be examination utilizing distinctive reenactment instruments at base station. Lakshmanan et al indicates the significance of IoT and their benefits and also threats handled by IoT [2].

Ayushi S jaiswal, Vaidehi Baporikar The data which is very difficult for humans to gather can be done by underwater robots. They are utilized broadly by mainstream researchers to think about sea submerged condition. Zigbee is a productive and viable remote system standard for remote control and checking applications. They displayed a reasonable and productive model of embedded remote information resource framework utilizing Zigbee which will be constrained by the PIC microcontroller [3].

In 2022, Manisha Koranga et.al discussed the use of Machine Learning Algorithms for water quality prediction for Nanital Lake, Uttarakhand. Analysed the use of machine learning algorithms and used eight regression algorithms and nine classification algorithms. Three algorithms Random Forest, SVM and Stochastic Gradient Descent comes out to be the most effective machine learning algorithms[4].Cheng-liang lai proposed utilizing picture preparing innovation for water quality checking framework. In which the effectively fabricated a water quality checking framework by using the picture handling framework and fluffy induction in auto perceiving the motion of fish [5].

The work done by Lakshmanan et al highlights the possible innovative communication methodologies and technologies for IoT [6]. Aravindan et al proposes the water quality management for real-time data [7].Vinod Raut et al uses an innovate wireless technique for monitoring the water quality [8]. A graphical UI was additionally actualized to show the information got and alert if any anomalies happened [9]. Jesudoss et al show how different types of sensors can be effectively implemented in safe driving [10].

III. SYSTEM ARCHITECTURE

Probably every individual is aware that water is one of the prime necessities for life of each living organism on the earth. The pH level and turbidity dimension of water assumes natural job in surveying the nature of water. Water quality assumes natural job in the medical problems of individuals, plants and living beings on the earth. Especially, the primary wellsprings of water are rivers, waterfalls, and lakes. Downpour water running over the grounds contains numerous impurities and polluting influences that might be dissolvable or insoluble. The primary point is to gauge the pH level and turbidity level in the drinking water just as in the sewage water from ventures that are crashed into the waterfalls and furthermore the water utilized for horticulture.

The goals of the framework are given beneath

- ✔ To sketch the analysis of the water quality monitoring system.
- ✔ To evaluate the pH and turbidity parameters in the real time environment using the sensors.

- ✔ To collect data from rural and less developed areas and store that data in the web page.

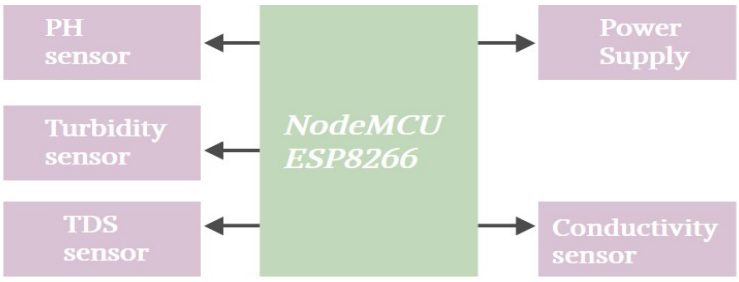
Thus the data collected is sent to the cloud with the help of wifi module present in the Arduino family microcontroller board. To exhibit the real time data on PC.

IV. HARDWARE SPECIFICATION

The pH,turbidity,conductivity and TDS parameters are used for checking the purity of water. These are calculated using the pH,turbidity,conductivity and TDS sensors which are connected to the Node MCU Microcontroller esp8266.

BLOCK DIAGRAM:

HARDWARE:



COMPONENTS:

A. PH SENSOR

PH sensor is used to determine whether the pH quality of the water is suitable for drinking or not. By measuring the pH level of water, the sensor provides information about its acidity or alkalinity, which can indicate its purity. The pH of most drinking-water lies within the range 6.5–8.5. The pH sensor data can contribute to a comprehensive analysis, enhancing the accuracy and reliability of the water purity testing system.

B. TURBIDITY SENSOR

Turbidity sensor is used to check the clarity of water, In this water purity tester, it can be used as a crucial input for machine learning algorithms to assess water quality. The sensor contains a light source ,typically an LED (Light Emitting Diode), and the *emitted light* is directed into the liquid sample, and measurement of the amount of light that is scattered by material in the water is observed. The higher the intensity of scattered light, the higher the turbidity.

C. CONDUCTIVITY SENSOR

In this project 2 pole conductivity sensor is employed. Here current is passed between the electrodes, and the conductivity of the solution is determined based on the amount of current that flows through the solution. It will imply that if conductivity is high then the rate of impurities is also high. By collecting conductivity data from the sensor, machine learning models can be trained to analyze patterns and relationships between conductivity levels and water purity. This approach allows for more precise identification of impurities or contaminants in the water, enabling better quality control and assurance.

D. TDS SENSOR

TDS sensor is used here to determine the total dissolved solids presence. By detecting the level of dissolved substances such as minerals, salts, metals, and other impurities, the TDS sensor helps assess the overall purity of the water. If the ppm lies between 50-300 then it is a cleaner water and if the values exceeds that range it is not safe to drink that water.

V. SOFTWARE SPECIFICATION

A. NodeMCU

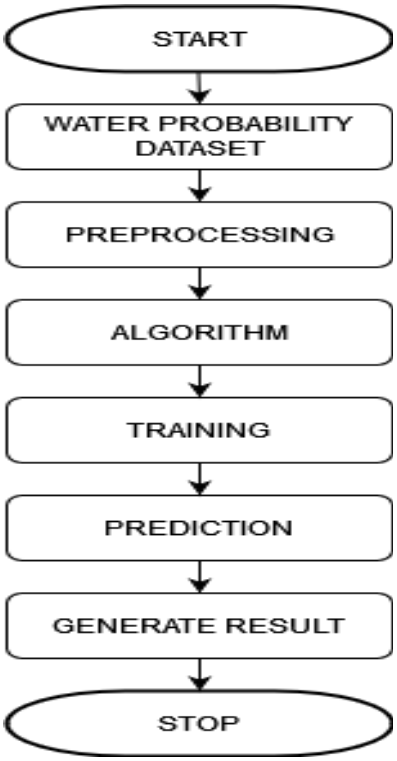
The PH sensor, Conductivity sensor, TDS sensor and Turbidity sensor will be connected to NodeMCU and the data observed through these sensors will be transmitted through the NodeMCU, for us to build the testing dataset.

B. ARDUINO IDE

Arduino IDE software is used to collect the data transmitted from NodeMCU and the datas will be converted to comma separated values format, so that the testing data can be given to the trained machine learning model to get evaluated.

C. DATA PREPROCESSING

To improve the water quality, pre-processing phase plays a vital role in data analysis. For the calculation of the Water Quality Index, the most significant factors are taken into consideration using PCA. For the system's superior accuracy, Data Normalization Techniques has been used.

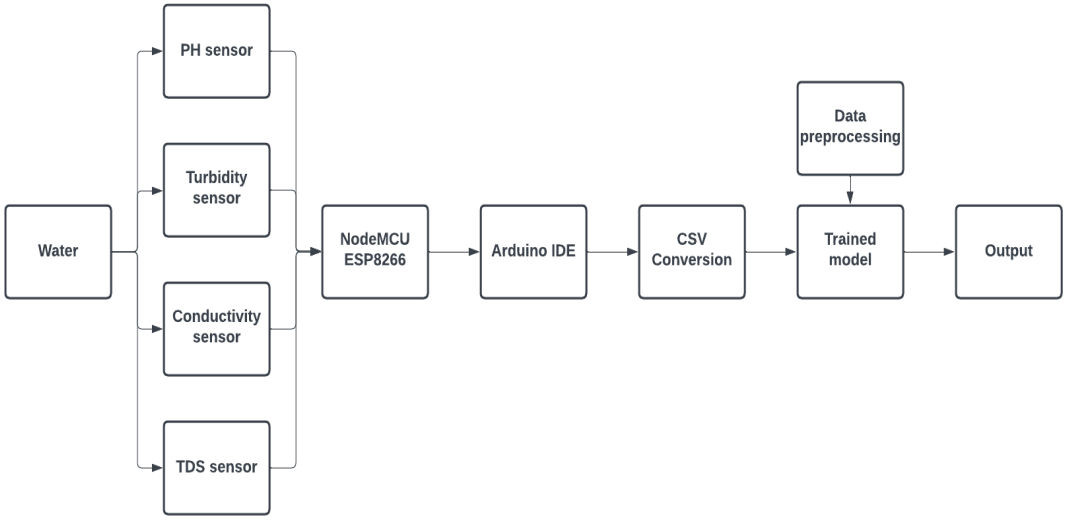


SOFTWARE(FLOW DIAGRAM)

VI. EXISTING SYSTEM

The current Water Quality observing framework include human towards looking at the water Quality, Testing. At present many measures are taken and innovative development has been introduced in water quality checking. These are finished by utilizing automated fish, laser bar and advanced camera. Likewise enquiry about the fish has been done by utilizing remote sensors. Notwithstanding observing the water quality, restricted work is

completed in applying AI strategy including the nature of water. The disadvantage of the current framework is that there is no completely robotized water Quality checking framework utilizing Sensors. Likewise framework does not have insight which takes into consideration dissecting the data for the forecast. These frameworks are worked for correspondence inside a little land zone.



FLOW DIAGRAM

A. PROPOSED SYSTEM

WORKING:

1. Hardware Setup: Connect the sensors such as pH sensor, turbidity sensor, conductivity sensor and TDS sensor to the NodeMCU ESP8266 board.

2. Arduino IDE Setup: Install the Arduino IDE and the necessary libraries.

3. Data Collection: Collect data from the sensors of water sample. Record the sensor readings along with the corresponding water purity level.

4. CSV Conversion: Format the data in a comma-separated format (CSV) by separating each value by comma.

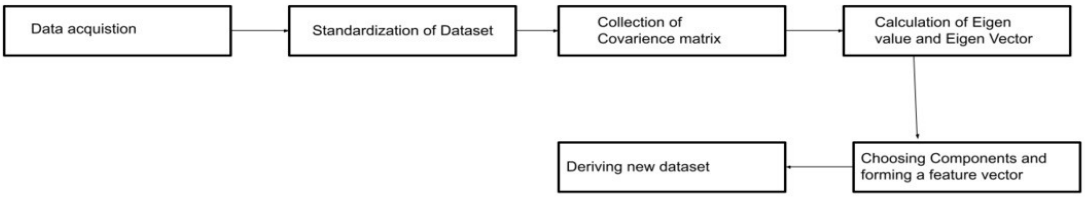
: 5. Machine Learning Model Training a machine learning model using the collected data and the extracted features. PCA(Principle Component Analysis) is used to reduce the dimensionality of the predefined dataset. Several Algorithms namely K-nearest neighbor(KNN),Support Vector Machine(SVM),Decision Tree are compared and noted that SVM has high accuracy.

6. Real-time Prediction: The Real-Time data will be given to the trained model as the Testing Data and result will be observed.

VII. ALGORITHMS

A.PCA(PRINCIPLE COMPONENT ANALYSIS):

PCA, or Principal Component Analysis, is a widely used technique in machine learning and statistics for dimensionality reduction. It's particularly helpful when dealing with high-dimensional datasets where the number of features (or variables) is large. PCA works by transforming the original variables into a new set of variables, which are linear combinations of the original ones, called principal components.



B. SUPPORT VECTOR MACHINE(SVM):

One of the most popular supervised learning algorithms is Support Vector Machine, it can be used for Regression and classification problems. Widley, it is used for Classification problems in Machine Learning.

Creation of the decision boundary (which is the best area or plane or line) that helps to sort n-dimensional data space into classes. This helps us to put the new query point in the accurate category in the future. Whenever there's a new query point, it is compared to the decision boundary and is classified accordingly. This is the main goal of Support Vector Machine.

C. DECISION TREE:

Decision tree is a supervised machine learning algorithm which can be used for both classification and regression.

It is mostly preferred to solve classification problems in which data has to be classified into different categories. It has a tree like structure with internal node acting as features, branches as decision rule and leaf nodes as the outcome.

The model is trained on our dataset using the decision tree approach. This type of classification problem is well suited for decision trees since they divide the feature space into regions according to the feature values.

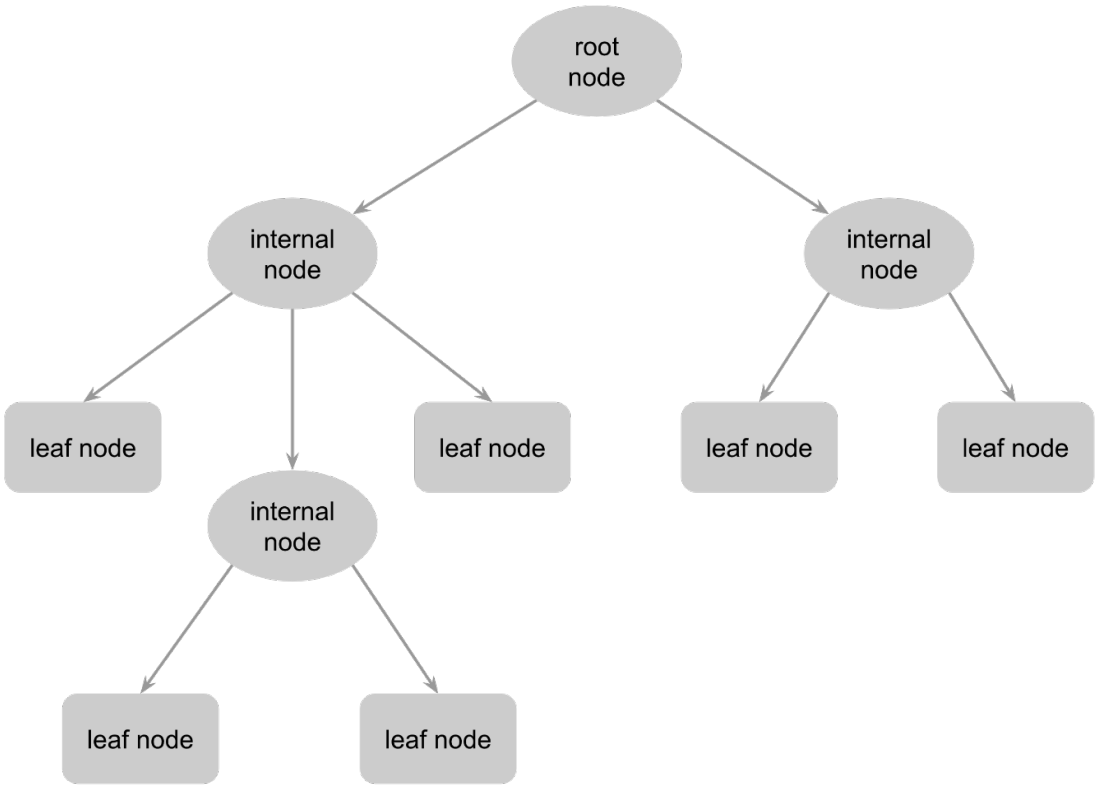
Utilizing metrics like accuracy, precision, recall, or F1 score to assess the trained model's performance.

1. Selecting the Best Feature: The algorithm evaluates different features and selects the one that best separates the data into classes or reduces variance (for regression).

2. Splitting the Data: It splits the data based on the selected feature into subsets.

3. Recursive Process: This splitting process is repeated on each subset until a stopping criterion is met, such as reaching a maximum tree depth, minimum number of samples in a node, or no further improvement can be made.

4. Leaf Nodes: Once the stopping criterion is met, the algorithm assigns a class label (or regression value) to each leaf node.



D. K- NEAREST NEIGHBOUR(KNN):

The K-Nearest Neighbors (KNN) algorithm is a simple and effective supervised learning algorithm used for classification and regression tasks. It classifies data points based on the majority vote of their nearest neighbors in the feature space. The 'K' represents the number of nearest neighbors considered for classification.

In the context of a water purity tester using the K-Nearest Neighbors (KNN) algorithm, the algorithm would classify water samples based on their similarity to neighboring samples in a feature space. For instance, features could include pH levels, dissolved oxygen content, and various chemical concentrations. KNN would classify a new water sample by examining the purity labels of its nearest neighbors, with the assumption that similar water samples tend to have similar purity levels.

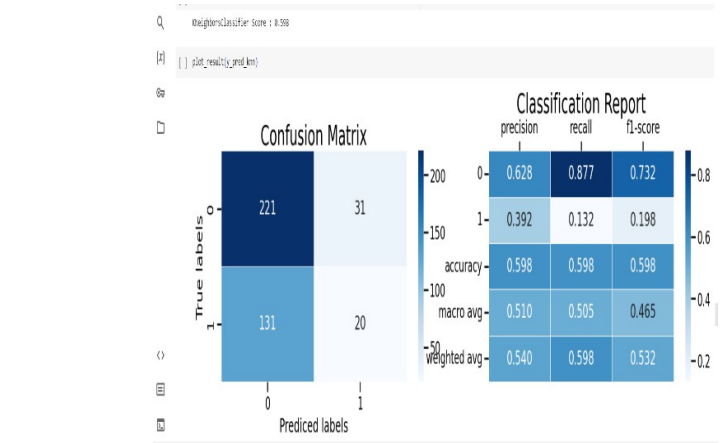
E. CONFUSION MATRIX:

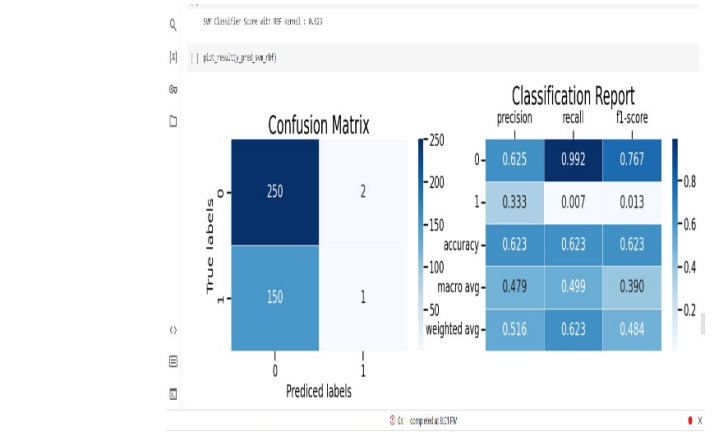
For classification problems confusion matrix is used on a wide scale. It is used for multiclass classification problems and binary classifications as well. Counts from predicted and actual values is represented by confusion matrices.

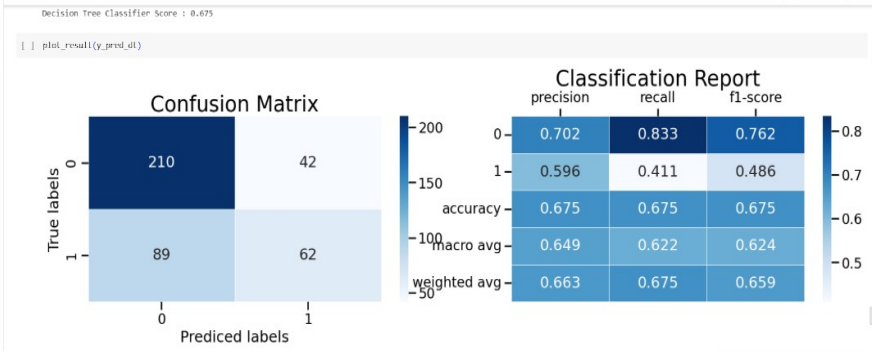
In confusion matrices True Negative is represented by "TN" it shows the number of negative examples which were labelled correctly. In the same way, True Positive is represented by "TP" it shows the number of positive samples which were labelled correctly. False Positive is represented by "FP" it shows the number of actual negative samples which were classified as positive. And False Negative is represented by "FN" it shows the number of actual positive samples classified as negative. For evaluation of WQI model we use Accuracy,Precision, Recall, Specificity, Mean Square Error, Sensitivity.

These statistical measurements are as mentioned below:

- Accuracy=TP+TN/TP+FP+FN+TN
- Precision=TP/TP+FP
- Recall=TP/TP+FN
- F1-score=2*P*R/(P+R)







VIII. RESULT

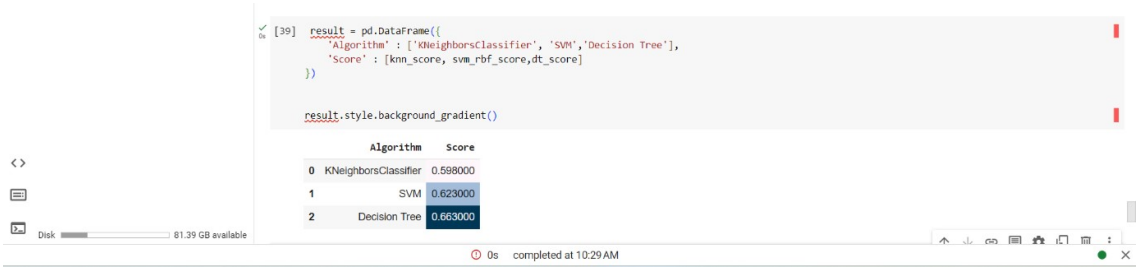
In this section, dataset used along with use of various Machine Learning algorithms for prediction is highlighted.

The dataset is used is from Kaggle. This dataset consists of water quality metrics for 3276 different water bodies. Key features used to tally the results are: pH value, hardness, solids, Chloramines, Sulphates, Organic carbon, Trihalomethanes, Turbidity, Potability. The results of the different machine learning algorithms, namely, SVM, Decision Tree, KNN are discussed based on parameters like accuracy, precision, recall and F1-score.

The main aspect taken into consideration is accuracy.

Accuracy=TP+TN/TP+FP+FN+TN

The results of the algorithms are mentioned below:



IX. CONCLUSION

The project concluded that using decision tree algorithm for water purity testing yielded the highest accuracy compared to PCA, SVM, and KNN. Therefore, decision tree was applied for the actual testing of water purity.

REFERENCES

[1] Mohammed Y Aalsalem, Wazir Zada Khan, Wajeb Gharibi, Nasrullah Armi "An intelligent oil and gas well monitoring system based on Internet of Things" International Conference on Radar, Antenna, Microwave, Electronics, and Telecommunications (ICRAMET),2017.

[2] Sayeda Islam Nahid, Mohammad Monirujjaman Khan " Toxic Gas Sensor and Temperature Monitoring in Industries using Internet of Things (IoT)" International Conference on Computer and Information Technology (ICCIT)2021

[3] S.Vivekanandan , Abhinav Koleti, M Devanand Autonomous industrial hazard monitoring robot with GSM integration International Conference on Engineering (NUiCONE)2013

[4] Meer Shadman Saeed, Nusrat Alim Design and Implementation of a Dual Mode Autonomous Gas Leakage Detecting Robot International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST)2019

[5] A.Sandeep Prabhakaran Mathan N Safety Robot for Flammable Gas and Fire Detection using Multisensor Technology International Conference on Smart Electronics and Communication (ICOSEC)2021.

[6] Ashutosh Mishra; Shiho Kim; N S Rajput" An Efficient Sensory System for Intelligent Gas Monitoring Accurate classification and precise quantification of gases/odors" International SoC Design Conference (ISOCC) 2020.

[7] Qiang Luo; Xiaoran Guo; Yahui Wang; Xufeng Wei "Design of wireless monitoring system for gas emergency repairing" Chinese Control and Decision Conference (CCDC) 2016.

[8] Mohammed Y Aalsalem; Wazir Zada Khan; Wajeb Gharibi; Nasrullah Armi "An intelligent oil and gas well monitoring system based on Internet of Things" International Conference on Radar, Antenna, Microwave, Electronics, and Telecommunications (ICRAMET) 2017.

[9] C.Nagarajan and M.Madheswaran - 'Experimental verification and stability state space analysis of CLL-T Series Parallel Resonant Converter' - Journal of ELECTRICAL ENGINEERING, Vol.63 (6), pp.365-372, Dec.2012.

[10] C.Nagarajan and M.Madheswaran - 'Performance Analysis of LCL-T Resonant Converter with Fuzzy/PID Using State Space Analysis'- Springer, Electrical Engineering, Vol.93 (3), pp.167-178, September 2011.

[11] C.Nagarajan and M.Madheswaran - 'Stability Analysis of Series Parallel Resonant Converter with Fuzzy Logic Controller Using State Space Techniques'- Taylor &; Francis, Electric Power Components and Systems, Vol.39 (8), pp.780-793, May 2011.

[12] C.Nagarajan and M.Madheswaran - 'Experimental Study and steady state stability analysis of CLL-T Series Parallel Resonant Converter with Fuzzy controller using State Space Analysis'- Iranian Journal of Electrical & Electronic Engineering, Vol.8 (3), pp.259-267, September 2012.

[13] Nagarajan C., Neelakrishnan G., Akila P., Fathima U., Sneha S. "Performance Analysis and Implementation of 89C51 Controller Based Solar Tracking System with Boost Converter" Journal of VLSI Design Tools & Technology. 2022; 12(2): 34–41p.

[14] C. Nagarajan, G.Neelakrishnan, R. Janani, S.Maithili, G. Ramya "Investigation on Fault Analysis for Power Transformers Using Adaptive Differential Relay" Asian Journal of Electrical Science, Vol.11 No.1, pp: 1-8, 2022.

[15] G.Neelakrishnan, K.Anandhakumar, A.Prathap, S.Prakash "Performance Estimation of cascaded h-bridge MLI for HEV using SVPWM" Suraj Punj Journal for Multidisciplinary Research, 2021, Volume 11, Issue 4, pp:750-756

[16] G.Neelakrishnan, S.N.Pruthika, P.T.Shalini, S.Soniya, "Perfromance Investigation of T-Source Inverter fed with Solar Cell" Suraj Punj Journal for Multidisciplinary Research, 2021, Volume 11, Issue 4, pp:744-749

[17] C.Nagarajan and M.Madheswaran, "Analysis and Simulation of LCL Series Resonant Full Bridge Converter Using PWM Technique with Load Independent Operation" has been presented in ICTES'08, a IEEE / IET International Conference organized by M.G.R.University, Chennai.Vol.no.1, pp.190-195, Dec.2007

[18] M Suganthi, N Ramesh, "Treatment of water using natural zeolite as membrane filter", Journal of Environmental Protection and Ecology, Volume 23, Issue 2, pp: 520-530,2022

[19] M Suganthi, N Ramesh, CT Sivakumar, K Vidhya, "Physiochemical Analysis of Ground Water used for Domestic needs in the Area of Perundurai in Erode District", International Research Journal of Multidisciplinary Technovation, pp: 630-635, 2019