River Water Quality Monitoring System

A PROJECT REPORT

Submitted by

DEENA DAYALAN K [211420104052]

DEEPAK S [211420104054]

ATMAKURI VENKATA NITHIN KUMAR [211420104032]

in partial fulfillment for the award of the degree
of
BACHELOR OF ENGINEERING

in

COMPUTER SCIENCE AND ENGINEERING

PANIMALAR ENGINEERING COLLEGE

(An Autonomous Institution, Affiliated to Anna University, Chennai)

APRIL 2024

PANIMALAR ENGINEERING COLLEGE

(An Autonomous Institution, Affiliated to Anna University, Chennai)

BONAFIDE CERTIFICATE

Certified that this project report **“RIVER WATER QUALITY MONITORING SYSTEM”** is the bonafide work “**DEENA DAYALAN K[211420104052],DEEPAK S[211420104054], ATMAKURI VENKATA NITHIN KUMAR [211420104032]”** who carried out the project work under my supervision.

Signature of the HOD with date
DR L.JABASHEELA M.E., Ph.D.,
PROFESSOR AND HOD

Department of Computer Science and Engineering,
Panimalar Engineering College,
Chennai - 123

Signature of the Supervisor with date
Mrs. S.T. SANTHANALAKSHMI,M.tech.,$\overline{Ph.D}$.,
SUPERVISOR AND ASSOCIATE PROFESSOR

Department of Computer Science and Engineering,
Panimalar Engineering College,
Chennai - 123

Certified that the above candidate(s) was examined in the End Semester Project Viva Voce Examination held on ....

INTERNAL EXAMINER **EXTERNAL EXAMINER**

DECLARATION BY THE STUDENT

We **DEENA DAYALAN K [211420104052],DEEPAK S[211420104054], ATMAKURI VENKATA NITHIN KUMAR [211420104032]** hereby declare that this project report titled **"RIVER WATER QUALITY MONITORING SYSTEM",** under the guidance of **Mrs. S.T. SANTHANALAKSHMI, M.Tech., $\overline{\boldsymbol{Ph.D}}$., as** the original work done by us and we have not plagiarized or submitted to any other degree in any university by us.

DEENA DAYALAN K[211420104052]

DEEPAK S[211420104054]

ATMAKURI VENKATA NITHIN KUMAR [211420104032]

ACKNOWLEDGEMENT

Our profound gratitude is directed towards our esteemed Secretary and Correspondent, **Dr.P.CHINNADURAI, M.A., Ph.D**., for his fervent encouragement. His inspirational support proved instrumental in galvanizing our efforts, ultimately contributing significantly to the successful completion of this project

We want to express our deep gratitude to our Directors, **Tmt. C.** VIJAYARAJESWARI, Dr. C. SAKTHI KUMAR, M.E., Ph.D., and Dr.**SARANYASREE SAKTHI KUMAR, B.E., M.B.A., Ph.D.,** for graciously affording us the essential resources and facilities for undertaking of this project.

Our gratitude is also extended to our Principal, **Dr. K. MANI, M.E., Ph.D.,** whose facilitation proved pivotal in the successful completion of this project.

We express our heartfelt thanks to **Dr. L. JABASHEELA, M.E., Ph.D.,** Head of the Department of Computer Science and Engineering, for granting the necessary facilities that contributed to the timely and successful completion of project.

We would like to express our sincere thanks to **Dr. N. Pughazendi, M.E, Ph.D.,** and **Mrs. S.T. SANTHANALAKSHMI, M.Tech,** $\overline{\boldsymbol{Ph.D}}$**.,** all the faculty members of the Department of CSE for their unwavering support for the successful completion of the project.

DEENA DAYALAN K[211420104052]

DEEPAK S[211420104054]

ATMAKURI VENKATA NITHIN KUMAR [211420104032]

ABSTRACT

Potability is a numerical representation of water quality, and this study focuses on quantifying potability using machine learning approaches. Potability is a statistic used to assess whether water is suitable for various uses and consumption. The research makes use of a number of water quality characteristics, such as conductivity, organic carbon, pH, hardness, particulates, chloramines, sulphate,and trihalomethanes, to assess the overall quality of the water. When combined, these parameters create a feature vector that the machine learning algorithms use as input data. Three classification techniques are used in the paper: K-Nearest Neighbour (KNN), Decision Tree (DT), and Artificial Neural Network (ANN). Based on the feature vector of water quality metrics, these techniques are used to estimate the water quality class. The goal of the project is to precisely forecast water potabilityusing machine learning techniques, which will support efforts to analyse and maintain water quality. Two datasets are used in the experiments: one is a synthetic dataset created at random using water quality characteristics, and the other is a real dataset that includes data collected from various sites around Andhra Pradesh. With this method, the machine learning models may be thoroughly tested andvalidated on a variety of datasets. The studies' findings show that the KNN classifier performs better than the other classifiers in predicting the potability of water. This implies that the KNN algorithm works especially well for correctly classifying water quality by identifying patterns and correlations within the feature vector.The results demonstrate how well machine learning techniques can predict the potability of water, underscoring its potential use in environmental monitoring and water resource management. Data mining, classification, potability, and water quality parameters are some of the important index phrases related to this study. These phrases sum up the main ideas of the study, whichfocuses on utilizing machine learning methods to evaluate water quality data and categorize water potability according to pertinent factors.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1
INTRODUCTON

1.1 OVERVIEW

Machine learning is a very hot topic for many key reasons, and because it provides the ability automatically obtain deep insights, recognize unknown patterns, and create high performing predictive models from data, all without requiring explicit programming instructions. The high level understanding is critical if ever involved in a decision-makingprocess surrounding the usage of machine learning, how it can help achieve business andproject goals, which machine learning techniques to use, potential pitfalls, and how to interpret the results. The most common machine learning tasks that one may come across while trying to solve a machine learning problem. Under each task are also listed a set ofmachine learning methods that could be used to resolve these tasks. Please feel free to comment/suggest if I missed mentioning one or more important points. Artificial intelligence, with its versatile applications, has become integral across diverse industries and professions. In image and speech recognition, AI algorithms analyze and interpret visual and auditory data, enabling applications ranging from facial recognition systems to virtual assistants like Siri and Alexa. In healthcare, AI aids in medical diagnosis by analyzing patient data to identify patterns and predict diseases with high accuracy. It is also employed in predictive analytics and classification tasks, assisting in risk assessment, customer segmentation, and fraud detection. Furthermore, AI algorithms excel in learning associations between data points, facilitating personalized recommendations in e-commerce and content streaming platforms. In finance, AI drivesstatistical arbitrage strategies by analyzing market trends and making high-frequency trading decisions. Moreover, AI enables data extraction and regression analysis, extracting valuable insights from vast datasets and predicting future outcomes in fields such as weather forecasting and supply chain management.

1.1 PROBLEM DEFINITION

The problem addressed by the River Water Quality Monitoring System using machine learning is the need for an effective and efficient method to monitor and assess the qualityof river water. Furthermore, with increasing pollution and environmental degradation, there is a growing demand for proactive measures to identify and mitigate potential risksto river ecosystems and public health. Thus, the problem entails developing a system capable of continuously monitoring key water quality parameters, analyzing large volumes of data and leveraging machine learning techniques to predict trends, classify water quality levels. The system should be scalable, adaptable, and user-friendly, cateringto the needs of various stakeholders, including environmental agencies, policymakers, researchers, and local communities, to support informed decision-making and effective management of river water resources.

1.2 OBJECTIVE

The River Water Quality Monitoring System employs machine learning to provide a flexible and all-encompassing framework for ongoing river water quality monitoring andevaluation. The system attempts to create precise prediction by combining real-time sensor data with past records and ambient parameters. These models will identify possible contamination events and threats to the environment in addition to classifying water quality levels and predicting future changes in water quality. The system would give stakeholders access to decision support tools and visualizations, facilitating well- informed choices on river water management and conservation initiatives. This will encourage cooperation and stakeholder participation in order to efficiently solve local water quality challenges. The ultimate goal of the River Water Quality Monitoring System is to improve the sustainability and protection of freshwater ecosystems by applying cutting-edge machine learning methods and tools.

CHAPTER 2

LITERATURE REVIEW

Toward Design of Internet of Things and Machine Learning-Enabled Frameworks for Analysis and Prediction of Water Quality[1] The degradation of water quality has becomea critical concern worldwide, necessitating innovative approaches for monitoring and predicting water quality. The IoT-enabled framework comprises four modules: sensing, coordinator, data processing, and decision. The IoT framework is equipped with temperature, pH, turbidity, and Total Dissolved Solids (TDS) sensors to collect the datafrom Rohri Canal, SBA, Pakistan. The acquired data is preprocessed and then analyzed using machine learning models to predict the Water Quality Index (WQI) and Water Quality Class (WQC).

A Novel Hybrid Model to Predict Dissolved Oxygen for Efficient Water Quality in Intensive Aquaculture[2] Dissolved oxygen content is a key indicator of water quality inaquaculture environment. Because of its nonlinearity, dynamics, and complexity, whichmakes traditional methods face challenges in the accuracy and speed of dissolved oxygencontent prediction. As a solution to these issues, this study introduces a hybrid model consisting of the Light Gradient Boosting Machine (LightGBM) and the Bidirectional Simple Recurrent Unit (BiSRU).

A Low-Complexity Machine Learning Nitrate Loss Predictive Model– Towards Proactive Farm Management in a Networked Catchment[3] With the advent of wireless sensor networks, the ability to predict nutrient-rich discharges, using on-node predictionmodels, offers huge potential for enabling real-time water reuse and management withinan agriculturally dominated catchment. Existing discharge models use multiple parameters and large historical data which are difficult to extract and this, coupled with constraints on network nodes (battery life, computing power, and sensor availability), makes it necessary to develop low-dimensional models.

Development of Chemical Oxygen on Demand (COD) Soft Sensor Using Edge Intelligence [4] compares the results of a few best performing ML algorithms like Multiple Linear Regression, Multilayer Perceptron, Support Vector Machine, Random Forest and K-Nearest Neighbour (KNN). Where, KNN technique proves to be the most efficient one in the prediction of COD in terms of response time and other performance matrices like R, R2, MSE, MAE, and RMSE. Finally, the preferred model

(KNN) with the IoT setup is deployed and tested at the Sewage Treatment Plant (STP) outlet of the authors' institute to verify the accuracy of the COD in real-time. Where, KNN technique proves to be the most efficient one in the prediction of COD in terms of response time and other performance matrices like R, R2, MSE, MAE, and RMSE. Finally, the preferred model (KNN) with the IoT setup is deployed and tested at the Sewage Treatment Plant (STP) outlet of the authors' institute to verify the accuracy of the COD in real-time.

Hybrid Machine Learning Ensemble Techniques for Modeling Dissolved Oxygen Concentration[5] The modeling performance was assessed using the statistical measuresof NashSutcliffe coefficient efficiency (NSE), Willmott's index of agreement (WI), rootmean square error (RMSE), mean absolute error (MAE) and mean square error (MSE) and correlation coefficient (CC). The results of the single AI-based models demonstratedthat HW (M3) served as the best model for predicting DO concentration. For ensemble results, BPNN-E (WI = 0.9764) was superior to the other three ET with average decreasedof more than 2% with regards to MAE. Investigation on the hybrid RF ensemble demonstrated the reliable accuracy for all the hybrid models with better predictive skill shown by the HW-RF (CC = 0.981) ensemble.

A Method Based on Improved Ant Colony Algorithm Feature Selection Combined WithGA-SVR Model for Predicting Chlorophyll-a Concentration in Ulansuhai Lake[6] Chlorophyll-a (Chl-a) is an important parameter of water bodies, but due to the complexity of optics in water bodies, it is currently difficult to accurately predict Chl-a concentration in water bodies by traditional methods. An adaptive ant colony exhaustive optimization (A-ACEO) algorithm is proposed for feature selection and combined with a novel intelligent algorithm of optimizing support vector regression (SVR) by genetic algorithm (GA) for prediction of Chl-a concentration.

Crop Yield Prediction Using Deep Reinforcement Learning Model for Sustainable Agrarian Applications[7] The proposed work constructs a Deep Recurrent Q-Network model which is a Recurrent Neural Network deep learning algorithm over the Q-Learning reinforcement learning algorithm to forecast the crop yield. The sequentially stacked layers of Recurrent Neural network is fed by the data parameters. The Q- learning network constructs a crop yield prediction environment based on the input parameters. Alinear layer maps the Recurrent Neural Network output values to the Q-values. The reinforcement learning agent incorporates a combination of parametric features with thethreshold that assist in predicting crop yield. Finally, the agent receives an aggregate score for the actions performed by minimizing the error and maximizing the forecast accuracy. The proposed model efficiently predicts the crop yield outperforming existing models by preserving the original data distribution with an accuracy of 93.7%.

Iterative Learning for Reliable Link Adaptation in the Internet of Underwater Things[8]This work therefore collectively takes into account multiple quality factors of communication at the same time by creating, analysing and validating the machine learning model to predict the most adequate communication parameters to solve the problem. The dataset of underwater wireless communication used in the learning modelswas obtained from measurements made in a real underwater environment near the Gulf of Incheon, South Korea, using a practical testbed designed and implemented by the authors. The estimated network throughput based on the communications parameters

Projected Water Levels and Identified Future Floods: A Comparative Analysis for Mahaweli River, Sri Lanka[9] The correlation coefficient of each algorithm's predictions was 0.9330, 0.9120, 0.9133, 0.8915, 0.6811, 0.6811, and 0.6734 for the Cascaded- ANFIS, LSTM, GRU, RNN, Linear, Ridge, and Lasso regression models respectively. Hence, this study concludes that the proposed algorithm is 21% more accurate than the second-best LSTM algorithm. In addition, Shared Socio-economic Pathways (SSP2-4.5 and SSP5-8.5 scenarios) were used to generate future rainfalls, forecast the near-future and mid-future water levels, and identify potential flood events. The future forecasting results indicate a decrease in flood events and magnitudes in both SSP2-4.5 and SSP5-8.5 scenarios. Furthermore, the SSP5-8.5 scenario shows drought weather from May to August yearly. The results of this study can effectively be used to manage and control water resources and mitigate flood damages.

Prediction of Dissolved Oxygen Content in Aquaculture Based on Clustering and Improved ELM parameter index[10] The prediction of dissolved oxygen can reduce theoperation cost of aquatic product management to a certain extent. In this paper, a hybridIn the aquaculture industry, dissolved oxygen is an important water quality method is proposed to predict the change of dissolved oxygen from the perspective of time series in aquaculture, which based on k-means clustering and improved Softplus extreme learning machine (SELM) with particle swarm optimization (PSO).

CHAPTER 3

THEORETICAL BACKGROUND

3.1 EXISTING SYSTEM

These systems typically leverage historical water quality data along with relevant environmental parameters to develop predictive models. Commonly employed machine learning algorithms include regression models, decision trees, support vector machines, and more recently, deep learning techniques like neural networks. These models are trained on datasets containing information on factors such as temperature, pH levels, dissolved oxygen, and pollutant concentrations. The predictive capabilities of these systems are used to forecast water quality metrics, aiding in the early detection of potential issues such as contamination events or deteriorating water conditions. Furthermore, some systems integrate real-time monitoring sensors to continuously update the model and improve prediction accuracy. Continuous advancements in machine learning and the increasing availability of water quality data contribute to the ongoing development and enhancement of such systems, playing a crucial role in water resource management and environmental monitoring. For the latest information, it is recommended to explore recent publications, research articles, or platforms dedicated towater quality prediction and monitoring.

3.1 PROPOSED METHODOLOGY

The proposed system is intended to determine potability. It is divided into two phases, one for training and the other for testing. The following procedures are carried out in bothsections. Data on training pH and hardness testing data Solids, chloramines, sulphate, conductivity, organic carbon, trihalomethanes, turbidity, and potability are all terms thatcan be used to describe something. The data set was chosen as follows: The collection ofessential parameters that affect water quality, identification of the number of data samples, and definition of the class labels for each data sample present in the data are all factors go into selecting the water quality data set, which is a prerequisite to model construction. Ten indicator parameters make up the data sets used in this study. pH valueand hardness are examples of these factors. Solids, chloramines, sulfate, conductivity, organic carbon, trihalomethanes, turbidity, and potability are all terms that can be used to describe the properties of a substance. The proposed approach, however, is not constrained by the number of parameters or the selection of parameters. A k-fold cross validation technique is employed to set the learning and testing framework in this study,corresponding to each data sample in the data set.

3.2 FEASIBILTY STUDY

A feasibility study is an evaluation and analysis of a project's potential based on thoroughinvestigation and research to provide decision-makers complete confidence. The goal offeasibility studies is to logically and objectively identify the advantages and disadvantages of a potential project or existing business, as well as the possibilities and dangers presented by the surrounding environment, the resources required to execute theplan, and, ultimately, the likelihood of success. The most important things to think aboutare: There are three levels of feasibility:1. economic,2. technical

3.2.1 ECONOMIC FEASIBILITY

Economic feasibility is a crucial aspect of a feasibility study that assesses the financial viability and potential economic benefits of a proposed project. It involves analysing thecosts associated with the project, as well as estimating the potential returns on investment.The primary goal is to determine whether the project is financially feasible and economically justifiable.

Potential Revenue Streams:

Purpose: To explore avenues for offsetting development and operational costs and achieving financial sustainability.

Return on Investment (ROI) Analysis:

Purpose: To assess the financial attractiveness of the project and its potential impact onhealthcare outcomes.

Cost-Benefit Analysis (CBA):

Purpose: To provide decision-makers with a clear understanding of the economic valueand feasibility of the project.

Payback Period Analysis:

Purpose: To assess the project's risk and provide insights into its financial payback period.

3.2.2 TECHNICAL FEASIBILITY

River water monitoring system can be a vital tool for environmental management, ensuring the health and safety of both ecosystems and human populations that rely on river resources. Deploying various sensors to monitor water quality parameters such as pH, dissolved oxygen, turbidity, temperature, conductivity, and nutrient levels. These sensors should be reliable, accurate, and capable of withstanding outdoor conditions. Establishing a robust communication network to transmit real-time data from remote monitoring locations to a central database or control center. This could involve wireless technologies such as cellular networks, satellite communication, or IoT protocols like LoRaWAN. Providing continuous power to monitoring stations, which may be located in remote or inaccessible areas. Options include solar panels, wind turbines, or battery backup systems. Designing the system to accommodate future expansion or modification as monitoring needs evolve. This may involve using modular components and open- source technologies that can be easily upgraded or replaced. The technical feasibility considerations, a river water monitoring system can provide valuable insights into the health of aquatic ecosystems, support early warning systems for pollution events, and facilitate evidence-based decision-making for sustainable water resource

3.3 IMPLEMENTATION ENVIRONMENT

The software requirements specification (SRS) represents a crucial milestone in system engineering, refining functions and performance attributes allocated to software post- analysis. It serves as a comprehensive guide, delineating system behavior, performance requirements, design constraints, and validation criteria. By encapsulating these aspects, the SRS lays the groundwork for development, facilitating clear communication betweenstakeholders and development teams, ensuring alignment with project goals, and guidingthe implementation of robust software solutions that effectively address stakeholder needs. Through precise understanding of requirements and adherence to validation criteria, development efforts are focused, risks are mitigated, and successful software delivery is assured.

Hardware Setup:

System	: Pentium IV 2.4 GHz
Hard Disk	: 40 GB
Floppy Drive	: 1.44 Mb
Monitor	: 15 VGA Colour
Mouse	: Logitech
Ram	: 512 Mb

Software Setup:

Operating system	: Windows 10
IDE	: anaconda navigator
Coding Language	: python

CHAPTER 4
SYSTEM DESIGN

4.1 FLOW DIAGRAM

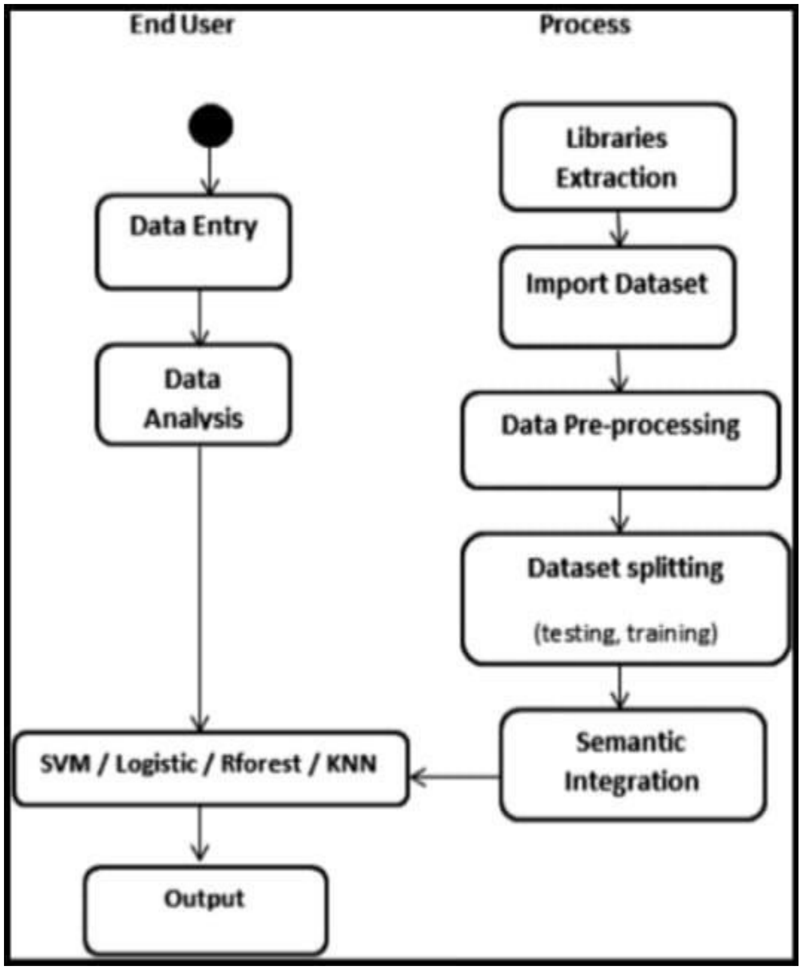


FIG 4.1.1 FLOWDIAGRAM

Flow diagram is a collective term for a diagram representing a flow or set of dynamic relationships in a system. The term flow diagram is also used as a synonym for flowchart,and sometimes as a counterpart of the flowchart.

4.2 UML DIAGRAMS

4.2.1 USECASE DIAGRAM

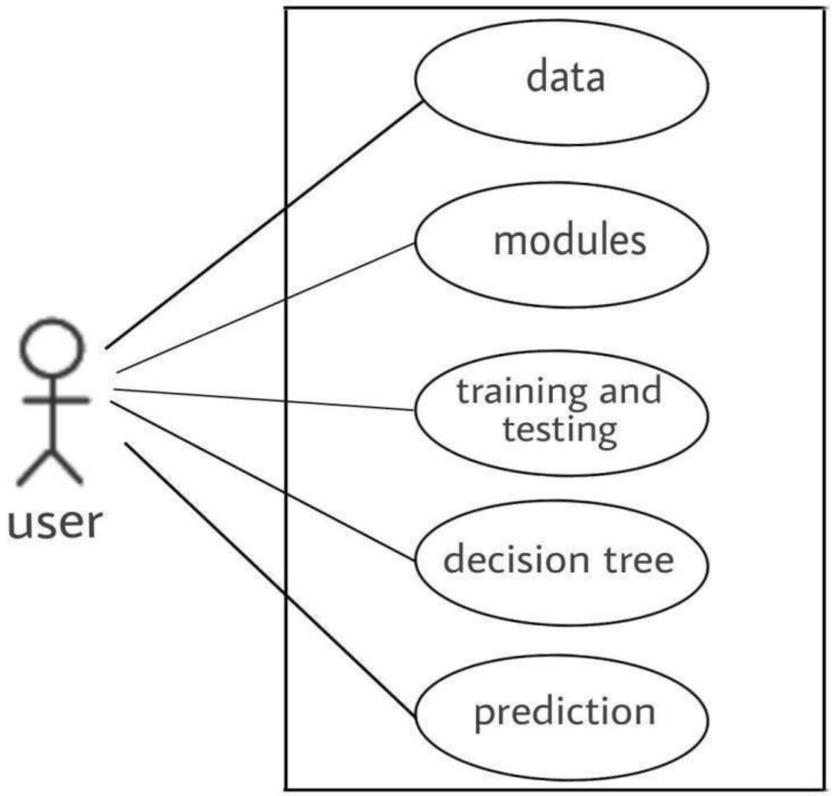


FIG 4.2.1 USECASE DIAGRAM

The block diagram shows the following steps involved in using a decision tree:

Data: This is the raw data that the decision tree will be trained on. In the context of theimage, the data could be a collection of water quality measurements, sensor readings, orother relevant information.

22Modules: This block represents the different modules that are used to train and test the decision tree. These modules could include a data pre-processing module, a training module, and a testing module.

Training and Testing: The training module is used to train the decision tree on the data.The testing module is used to test the performance of the decision tree on new data.

User: This block represents the user of the decision tree. The user can provide input to the decision tree, such as a new water sample, and the decision tree will make a prediction

Decision Tree: This is the core of the block diagram. The decision tree is a treelikestructure that consists of nodes and branches. Each node represents a decision that ismade based on the data. Each branch represents the possible outcomes of that decision. **Prediction:** This block represents the output of the decision tree. The prediction is thecategory that the decision tree has classified the new data point as belonging to. In thecontext of the image, the prediction would be whether or not the water sample is potable.

4.2.2 ACTIVITY DIAGRAM

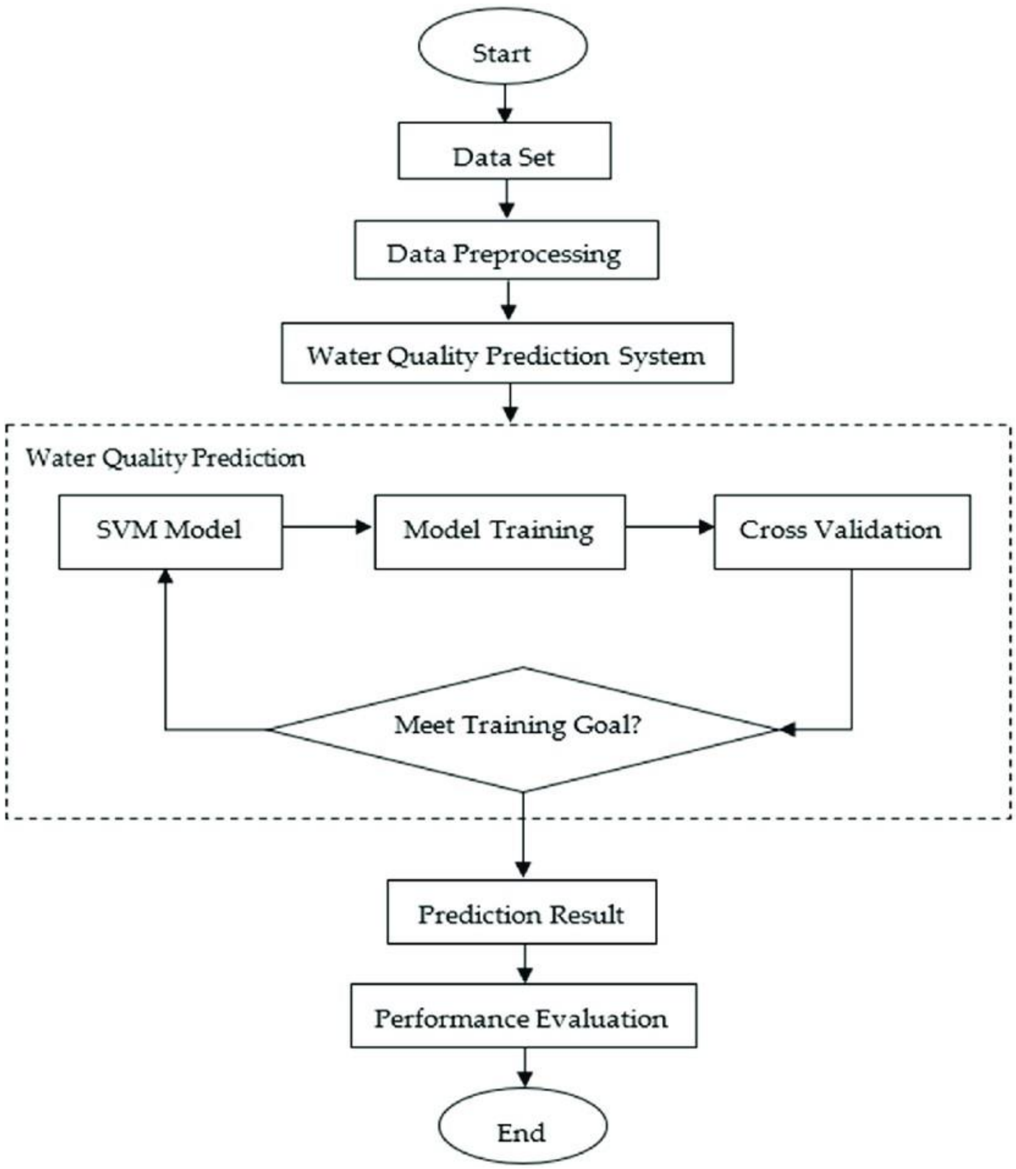


FIG 4.2.2 ACTIVITY DIAGRAM

The water quality prediction system embarks on its journey with the initiation of the data acquisition phase, where historical water quality data is meticulously collected to form a robust dataset for training the prediction model. This dataset encompasses a diverse range of parameters spanning physical attributes such as temperature, pH, conductivity, and turbidity, chemical properties including dissolved oxygen, nitrates, phosphates, and heavy metals, as well as biological factors like algae, bacteria, and protozoa. Following data acquisition, a thorough preprocessing stage ensues, aimed at refining the dataset for machine learning analysis. This involves addressing missing values, detecting and eliminating outliers, standardizing feature scales, and encoding categorical variables into numerical representations, ensuring the dataset's suitability for subsequent modeling.

With the dataset prepared, the system pivots its focus towards model training and evaluation. At the heart of the system lies the water quality prediction model, where various machine learning algorithms are considered, including Support Vector Machines (SVMs), Artificial Neural Networks (ANNs), Long Short-Term Memory (LSTM) networks, Random Forests, and XGBoost. The selection of the model hinges upon the specific problem context and characteristics of the data. Model training entails exposing the chosen model to the preprocessed data, enabling it to discern intricate relationships between input features and output targets. Subsequently, cross-validation techniques are employed to assess the model's performance on unseen data, thereby mitigating the risk of overfitting and ensuring its generalization capability across different scenarios.

4.2.3 SEQUENCE DIAGRAM

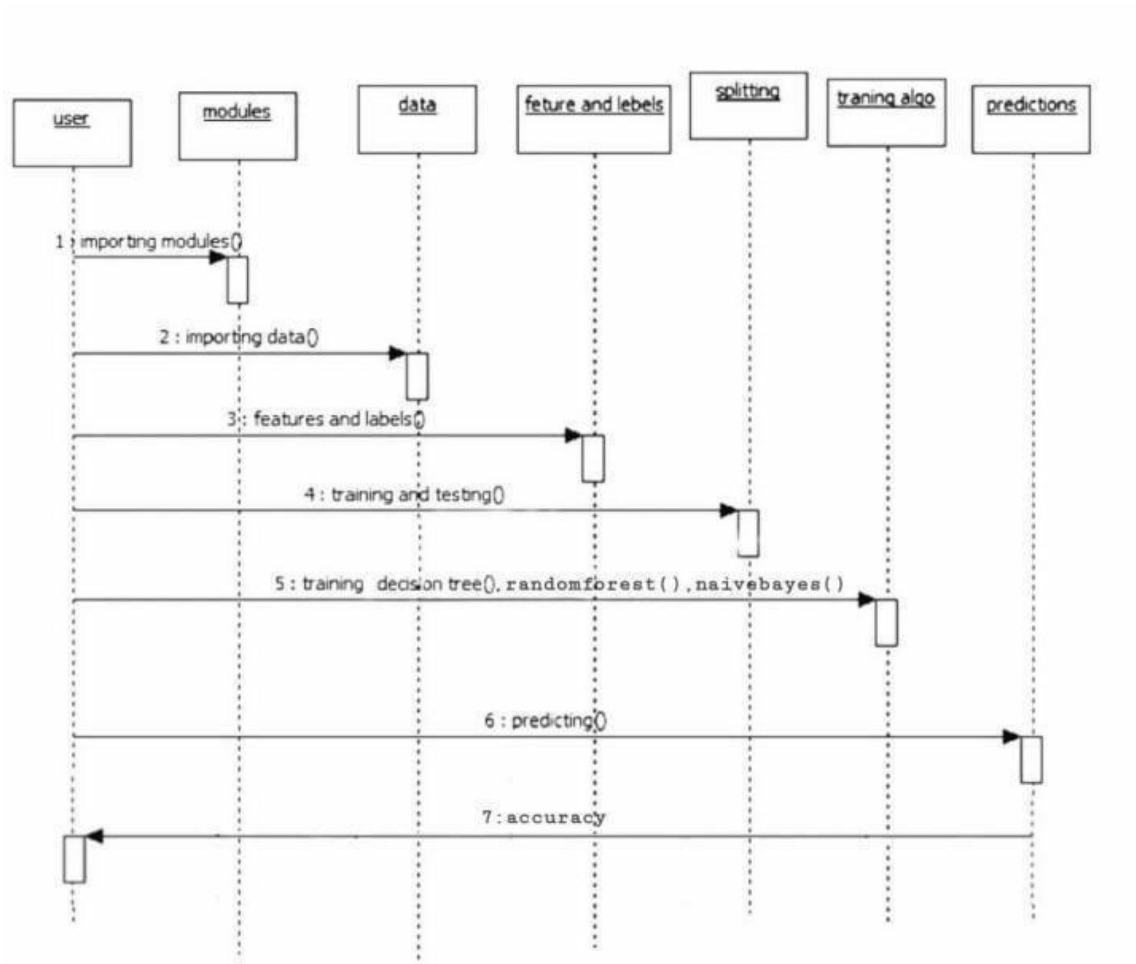


FIG 4.2.3 SEQUENCE DIAGRAM

The sequence diagram illustrates the comprehensive process of developing and implementing a water quality prediction system. It begins with the initiation of the system, followed by the importation of real-time or historical water quality data from various sources. This raw data then undergoes preprocessing, which involves cleaning and preparing it for input into the machine learning model. Tasks such as handling missing values, identifying and removing outliers, and encoding categorical variables ensure the data's quality and suitability for analysis. Moving forward, the model trainingand selection phase involves selecting relevant features from the preprocessed data and training a chosen machine learning model. This model may include Support Vector Machines, Artificial Neural Networks, LSTM networks, Random Forests, or XGBoost, depending on the specific requirements of the prediction task. Hyperparameter tuning optimizes the model's performance, and model evaluation on a validation set determines the best-performing model for prediction. In the prediction and evaluation stage, theselected model is deployed to predict future water quality parameters based on new data. Continuous performance monitoring using metrics like mean squared error or meanabsolute error ensures the model's reliability and accuracy over time. The iterative natureof the process allows for ongoing data collection, model updates, and prediction generation to inform decision-making in water quality management.

4.2.4 COMPONENT DIAGRAM

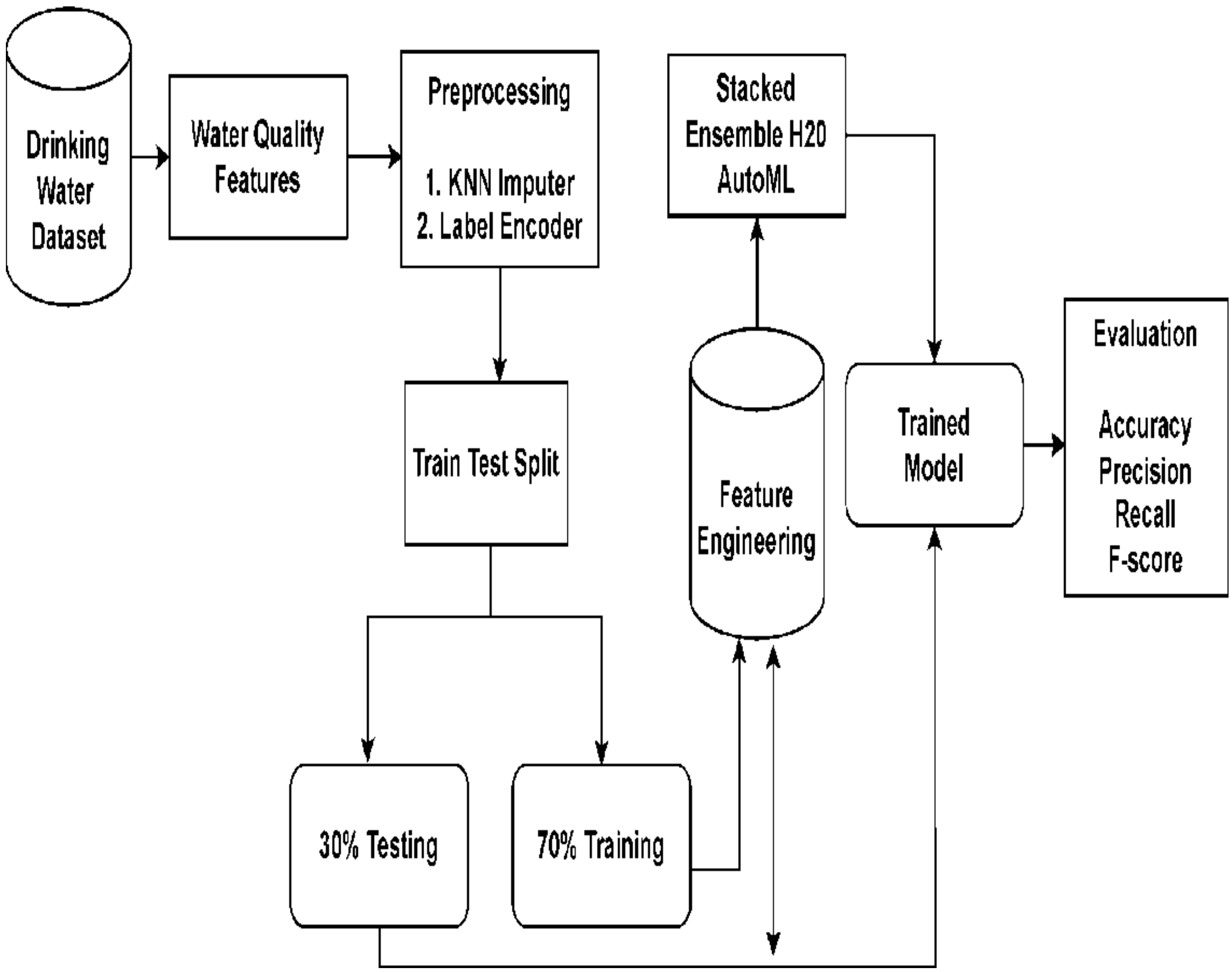


FIG 4.2.4 COMPONENT DIAGRAM

The diagram presents a streamlined depiction of a water quality prediction system, delineating the sequential progression of operations from data acquisition to prediction analysis. At the outset, the system acquires data from a comprehensive dataset containinginformation pertinent to drinking water quality, spanning various parameters such as physical, chemical, and biological attributes. This raw data then undergoes meticulous preprocessing, involving tasks like handling missing values, outlier removal, feature scaling, and encoding categorical variables into numerical representations, ensuring its suitability for subsequent analysis. 28Subsequently, the model training and ensemble creation phase employ sophisticated techniques to harness the predictive power of the data. The KNearest Neighbors (KNN) imputation method fills missing values with the average of neighboring values, while the label encoder transforms categorical variables into numerical formats compatible with machine learning models. A train-test split segregates the data for training and evaluation purposes, paving the way for the StackedEnsemble H2O AutoML platform. This core component autonomously constructs and fine-tunes multiple machine learning models, employing a stacked ensemble strategy to amalgamate predictions and enhance overall accuracy. Evaluation ensues to assess the performance of the trained model, leveraging metrics such as accuracy, precision, recall,and F1-score to gauge its effectiveness in generalizing to unseen data. Finally, the systemproceeds to prediction and analysis, utilizing the trained model to generate water qualitypredictions for new data points. Employing Explainable AI techniques facilitates a deeperunderstanding of the model's decision-making process, enhancing interpretability and trust. Ultimately, the system outputs a set of water quality predictions, poised to inform critical decision-making processes related to water resource management, pollution control, and public health safeguarding.

CHAPTER 5

SYSTEM ARCHITECTURE

5.1 ARCHITECTURE OVERVIEW

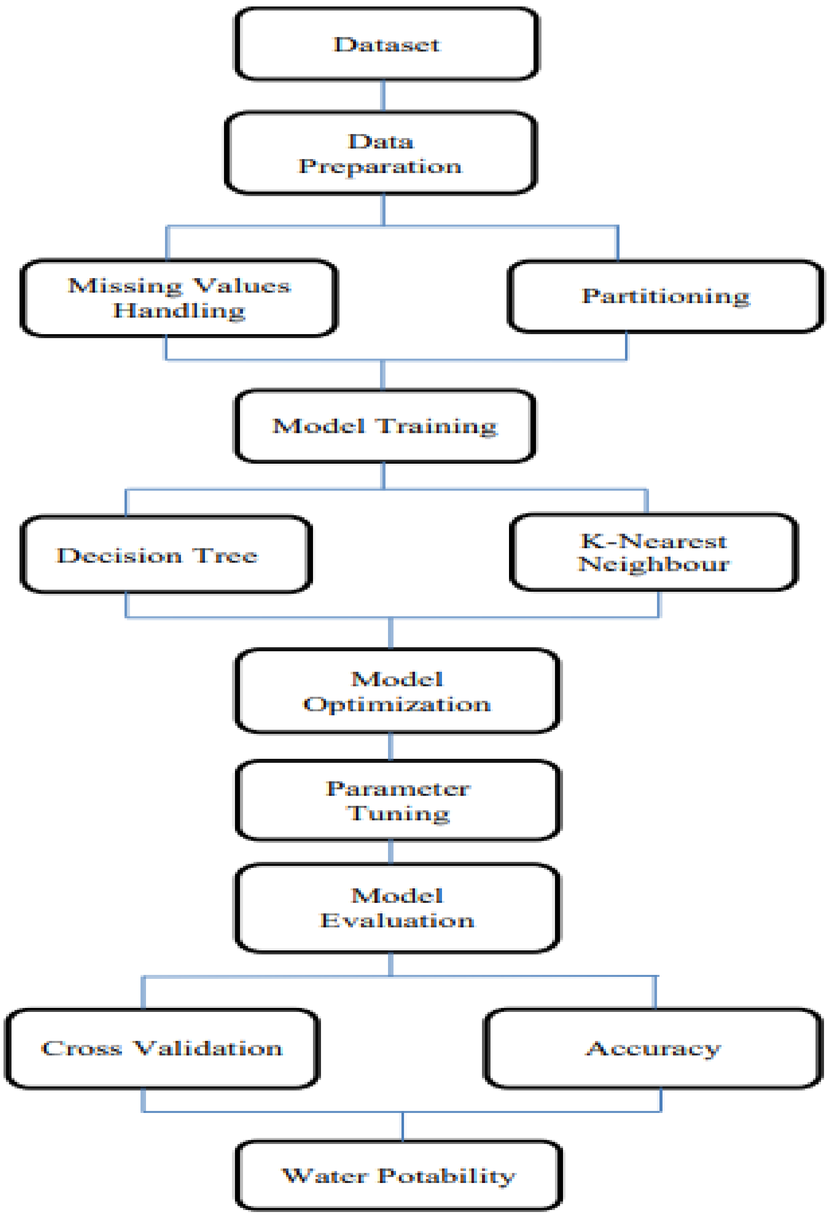


FIG 5.1.1 ARCHITECTURE DIAGRAM

Design is a multi- step that focuses on data structure software architecture, procedural details, procedure and interface among modules. The design procedure also decodes therequirements into presentation of software that can be accessed for excellence before coding begins. Computer software design change continuously as novel methods; improved analysis and border understanding evolved. Software proposal is at relatively primary stage in its revolution.Therefore, software design methodology lacks the depth, flexibility and quantitative nature that are usually associated with more conventional engineering disciplines. However, methods for software designs do exit, criteria for design qualities are existing and design notation can be applied.

DATASET:

A dataset serves as the foundational raw material for analysis, decision-making, and modeling across various domains, encompassing quantitative and qualitative data types organized into variables and attributes. It includes metadata detailing data descriptions, sources, and quality assessments, along with information on data structures, collection processes, and preprocessing steps. Represented in tabular, hierarchical, or temporal formats, datasets undergo cleaning, transformation, and normalization to prepare them for analysis. Stored in various formats and managed by data systems, datasets adhere to access controls and sharing agreements, ensuring validity, reliability, and ethical use in generating insights and driving informed decisions.

DATA PREPARATION:

In the data preparation step, the focus lies on cleaning and pre-processing the raw datasetto ensure its suitability for analysis or machine learning modeling. This encompasses addressing missing values through imputation or removal, identifying and handling outliers that could skew results, and formatting the data into a consistent and interpretableform. Techniques such as normalization, scaling, and encoding categorical variables are applied to ensure uniformity and compatibility with machine learning algorithms. The ultimate goal is to refine the dataset to a state where it can effectively serve as input for subsequent analytical or modeling tasks, enhancing the accuracy and robustness of the resulting insights or predictions.

FEATURE SELECTION:

In the feature selection phase for characterizing water quality, relevant features such as pH levels, temperature, dissolved oxygen, turbidity, and nutrient concentrations are identified from the dataset. This selection is based on their known significance in reflecting water quality parameters. Subsequently, the data undergoes thorough cleaningand preprocessing to handle missing values, outliers, and ensure uniformity across features. Techniques like imputation, outlier detection, and normalization are applied toaddress inconsistencies and prepare the dataset for further analysis. This meticulous process ensures that the selected features accurately represent the water quality characteristics and are suitable for subsequent modeling or analysis tasks.

MODEL TRAINING:

In the phase of model training, the prepared dataset undergoes the process of training a machine learning model to learn patterns and relationships within the data. Two potentialmodels depicted in the diagram are the decision tree and K-nearest neighbors (KNN) model. The decision tree model utilizes a hierarchical tree-like structure to make decisions based on feature splits, while the KNN model classifies data points based on the majority class of their nearest neighbors. The choice between these models depends on the characteristics of the data and the nature of the problem being addressed. Decisiontrees are suitable for tasks where interpretability and explainability are important, while KNN tends to perform well in classification tasks with nonlinear boundaries and instances where instances are not linearly separable. The selection of the appropriate model is crucial as it directly impacts the model's performance and its ability to generalizeto unseen data.

MODEL OPTIMIZATION

After training the machine learning model, the next step involves fine-tuning its parameters to optimize its performance, a process known as hyperparameter tuning. Hyperparameters are configuration settings external to the model that govern its learningprocess, such as learning rate, regularization strength, or tree depth. Through techniqueslike grid search, random search, or Bayesian optimization, various combinations of hyperparameters are systematically explored to find the configuration that maximizes themodel's performance metrics, such as accuracy, precision, or recall. Hyperparameter tuning aims to strike a balance between model complexity and generalization ability, ultimately enhancing the model's predictive capabilities and adaptability to new data.

VALIDATION:

In the validation phase, the trained model's performance is rigorously assessed using a separate dataset that was not utilized during the training process, ensuring its ability to generalize to new data. Techniques such as cross-validation or holdout validation are employed to evaluate the model's accuracy, precision, recall, or other relevant metrics. This validation step serves as a critical measure to ascertain the model's robustness and reliability in making predictions or classifications. Following successful validation, the trained model is implemented into a system interfacing with water quality sensors for real-time monitoring. This integration enables the model to continuously analyze incoming sensor data, providing timely insights and alerts about water quality conditions, thereby facilitating proactive management and decision-making in environmental monitoring applications.

MODEL EVALUATION

In the final stage of model evaluation, the performance of the trained model is assessed on a separate dataset that it hasn't encountered before, a process known as cross- validation. This step is essential for gauging the model's ability to generalize to new, unseen data and to mitigate overfitting. Cross-validation involves partitioning the datasetinto multiple subsets, typically through techniques like k-fold cross-validation or leave- one-out cross-validation, where the model is trained on a portion of the data and then tested on the remaining unseen data. Performance metrics such as accuracy, precision, recall, or F1 score are calculated to evaluate the model's effectiveness in makingpredictions. By iteratively repeating this process across different subsets, a robust assessment of the model's performance is obtained. This framework serves as a fundamental workflow applicable to various machine learning tasks, providing a systematic approach to building and assessing predictive models.

ALERT SYSTEM

To establish an alert mechanism for monitoring water quality deviations, a comprehensive system can be developed with real-time sensor integration to continuously monitor water parameters. When deviations from acceptable levels are detected, alerts are triggered automatically, notifying relevant authorities or users via various communication channels such as email, SMS, or mobile app notifications. This interface allows for easy navigation and interpretation of data, facilitating effective monitoring and timely response to ensure the safety of water resources. Integration with historical data and predictive models further enhances the system's capability to anticipate potential issues and proactively address them, thereby safeguarding water quality and public health.

5.2 ALGORITHM

5.2.1 ARTIFICAL NEURAL NETWORK

Artificial Neural Network Tutorial provides basic and advanced concepts of ANNs. OurArtificial Neural Network tutorial is developed for beginners as well as professions. Theterm “Artificial neural network” refers to a biologically inspired sub-field of artificial intelligence 23 odelled after the brain. An Artificial neural network is usually a computational network based on biological neural networks that construct the structure of the human brain. Similar to a human brain has neurons interconnected to each other, artificial neural networks also have neurons that are linked to each other in various layersof the networks. These neurons are known as nodes. Artificial neural network tutorial covers all the aspects related to the artificial neural network. In this tutorial, we will discuss ANNs, Adaptive resonance theory, Kohonen self-organizing map, Building blocks, unsupervised learning, Genetic algorithm, etc. The term “Artificial Neural Network” is derived from Biological neural networks that develop the structure of a human brain. Similar to the human brain that has neurons interconnected to one another,artificial neural networks also have neurons that are interconnected to one another in various layers of the networks. These neurons are known as nodes.

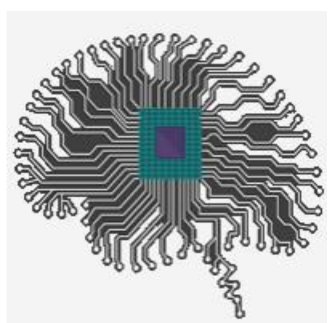


FIG 5.2.1 ARTIFICIAL NEURAL NETWORK

5.2.2 SUPPORT VECTOR MACHINE

The SVM algorithm using Python. Use the same dataset **user_data**, which we have usedin Logistic regression and KNN classification. The steps for using Support Vector Machine (SVM) algorithm in disease management:

Data collection:

Collect data on patients with the disease of interest, including information such as demographic data, medical history, symptoms, and lab results.

Data pre-processing:

Pre-process the data to ensure it is in a format that can be used by the SVM algorithm. This may involve data cleaning, data normalization, and feature engineering.

Data splitting:

Split the pre-processed data into a training set and a testing set. The training set will be used to train the SVM algorithm, and the testing set will be used to evaluate its performance.

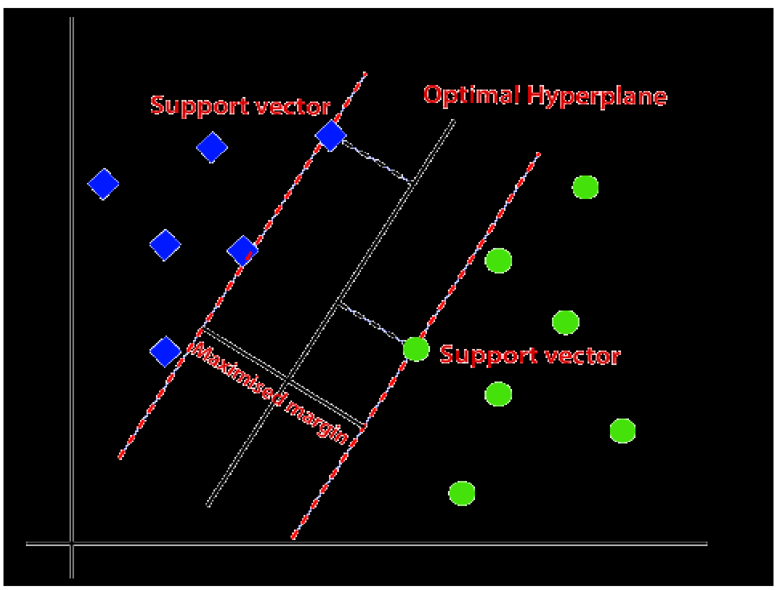
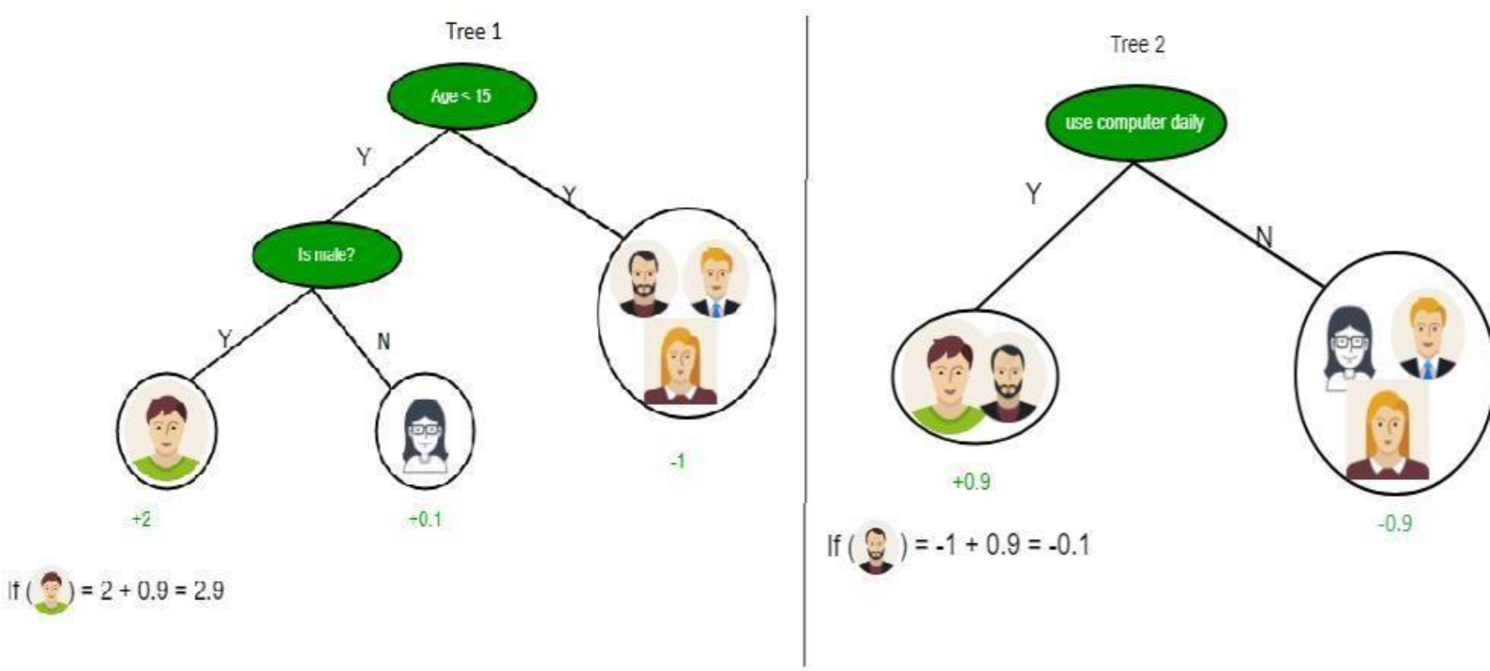


FIG 5.2.2 LINEAR SVM

5.2.3 DECISION TREE

The decision tree algorithm is a fundamental tool in supervised learning, applicable to both regression and classification problems. It employs a tree-like structure where each internal node represents a decision based on a feature attribute, and each leaf node corresponds to a class label or regression value. By iteratively partitioning the feature space based on the values of different attributes, decision trees efficiently capture complex decision boundaries and relationships within the data. This versatility allows decision trees to represent any boolean function on discrete attributes, making them widely applicable across various domains for tasks ranging from predicting outcomes toclassifying data.



5.3MODULE DESCRIPTION

Collecting Dataset

Data collection for machine learning involves gathering relevant data from various sources such as databases, APIs, or web scraping, based on the requirements of the problem domain. However, the collected dataset often contains noise, missing values, outliers, and other inconsistencies. Pre-processing the data is crucial to clean, transform,and format it into a suitable structure for the algorithm. This involves steps like handlingmissing values, removing duplicates, scaling features, encoding categorical variables, and detecting and handling outliers to obtain a refined dataset that enhances the algorithm's performance and generalization capabilities.

Data pre-processing

Data pre-processing is a vital step in preparing raw data for machine learning algorithms,typically comprising three main stages: data cleaning, transformation, and selection. Data cleaning involves identifying and rectifying errors, inconsistencies, and missing values within the dataset. Transformation includes standardizing or normalizing features, handling categorical variables through encoding techniques, and addressing skewness or distributional issues. Data selection entails choosing relevant features and eliminating redundant or noisy ones to improve model performance and efficiency. By systematically performing these pre-processing tasks, data scientists can enhance the quality and reliability of the dataset, leading to more accurate and robust machine learning models.

Data Cleaning

Data pre-processing encompasses several crucial tasks such as filling in missing values, smoothing noisy data, identifying or removing outliers, and resolving inconsistencies. Filling in missing values involves techniques like mean imputation, interpolation, or predictive modeling to estimate and substitute missing data points. Smoothing noisy data aims to reduce random fluctuations in the dataset by applying techniques like moving averages or kernel smoothing. Identifying or removing outliers helps in ensuring that abnormal data points do not unduly influence the model's behavior,often through statistical methods or domain-specific knowledge. Resolving inconsistencies involves detecting and rectifying discrepancies in the dataset to ensure itsintegrity and reliability for subsequent analysis and modeling tasks. Through theseconcerted efforts, data pre-processing enhances the quality, accuracy, and robustness of machine learning models.

Data transformation

May include smoothing, aggregation, generalization, transformation which improves thequality of the data.

Data selection Includes some methods or functions which allow us to select the usefuldata for our system.

Data Input Temperature

The water temperature in degrees Celsius. The pH level, representing the acidity or alkalinity of the water. A measure of the cloudiness or haziness of the water.The amount of oxygen dissolved in the water, measured in milligrams per liter (mg/L). Theability of the water to conduct an electric current, often measured in microsiemens percentimeter (µS/cm). The water quality parameter you want to predict or classify (e.g., "Good," "Moderate," "Excellent," "Poor")

Model Developement

Once you are satisfied with the performance of your model, deploy it in a realtime monitoring system. This could involve integrating it with sensors deployed in the riveror creating a web-based interface for users to access the predictions

CHAPTER 6
PERFORMANCE ANALYSIS

6.1 SYSTEM TRAINING

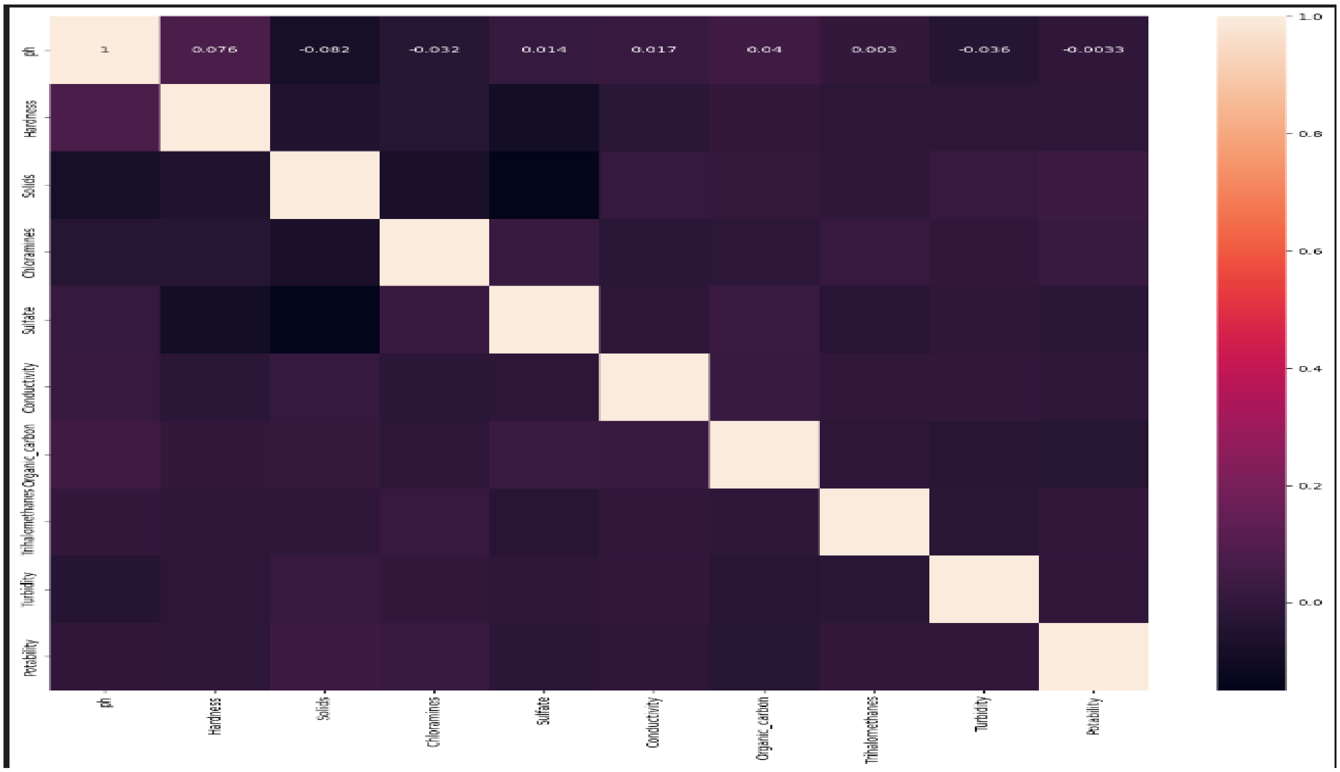


FIG 6.1.1 SCREENSHOT OF HEATMAP

A heatmap for a river water monitoring system could display various parameters across different sections of the river, providing a visual representation of water quality and environmental conditions.Parameters such as pH levels, dissolved oxygen concentration, temperature, turbidity, and pollutant levels could be color-coded on the heatmap, with different colors indicating different levels of each parameter. This visualization would allow stakeholders to quickly identify areas of concern or areas with optimal water quality, facilitating targeted interventions or further investigation where necessary.trends over time could be visualized by comparing heatmaps from different time periods, enabling long-term monitoring and assessment of changes in water quality and environmental health.

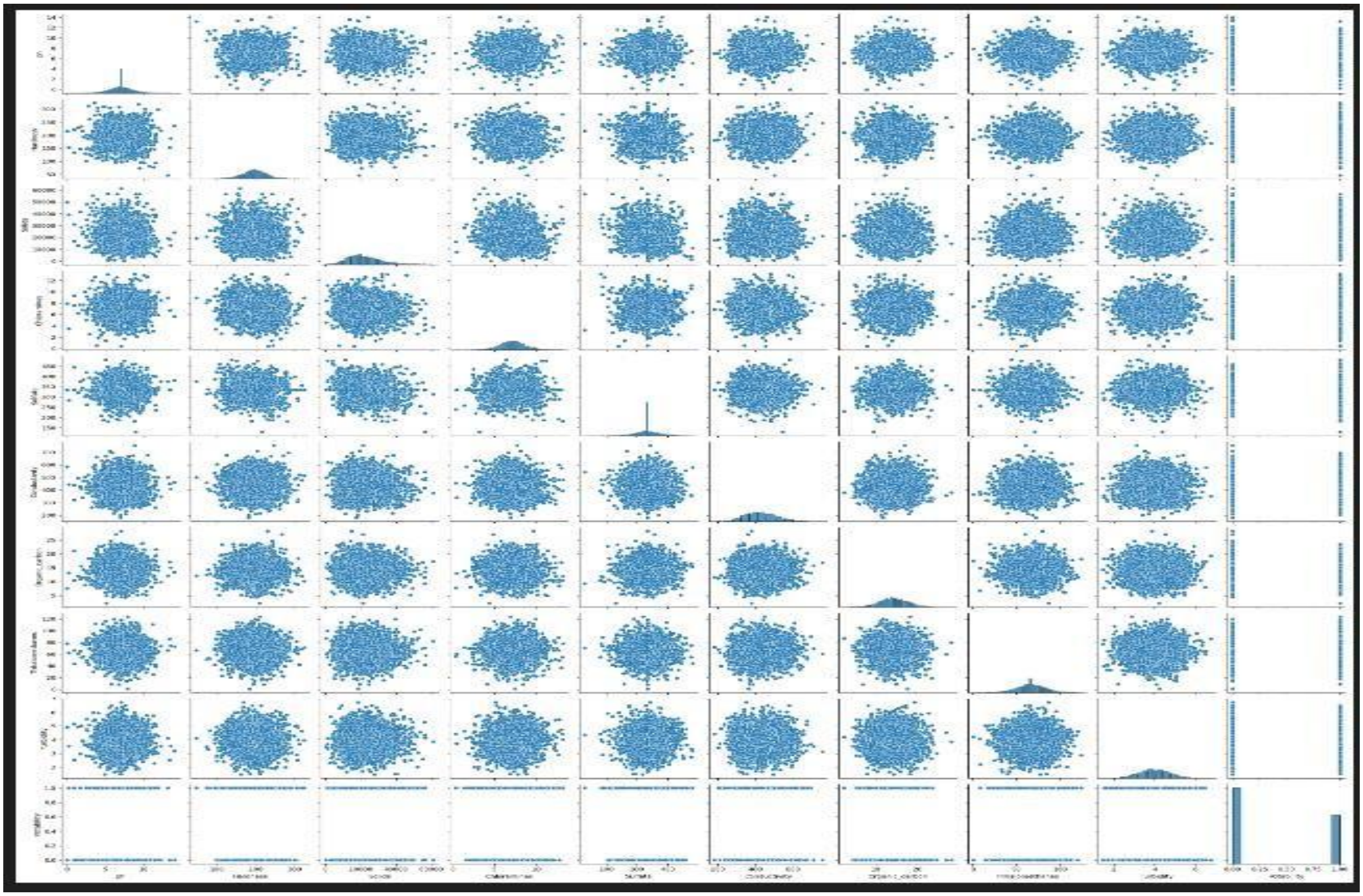


FIG 6.1.2 SCREENSHOT OF PLOT

Seaborn, a versatile Python library for statistical data visualization, can be effectively utilized in a riverwater monitoring system to generate insightful visualizations. Through Seaborn's high-level interface,stakeholders can create various plots such as line plots to observe temporal trends, scatter plots to analyze correlations between water quality parameters, and heatmaps to visualize spatial variations along the river. These visualizations provide valuable insights into water quality dynamics, aiding in decision-making processes and environmental management efforts.

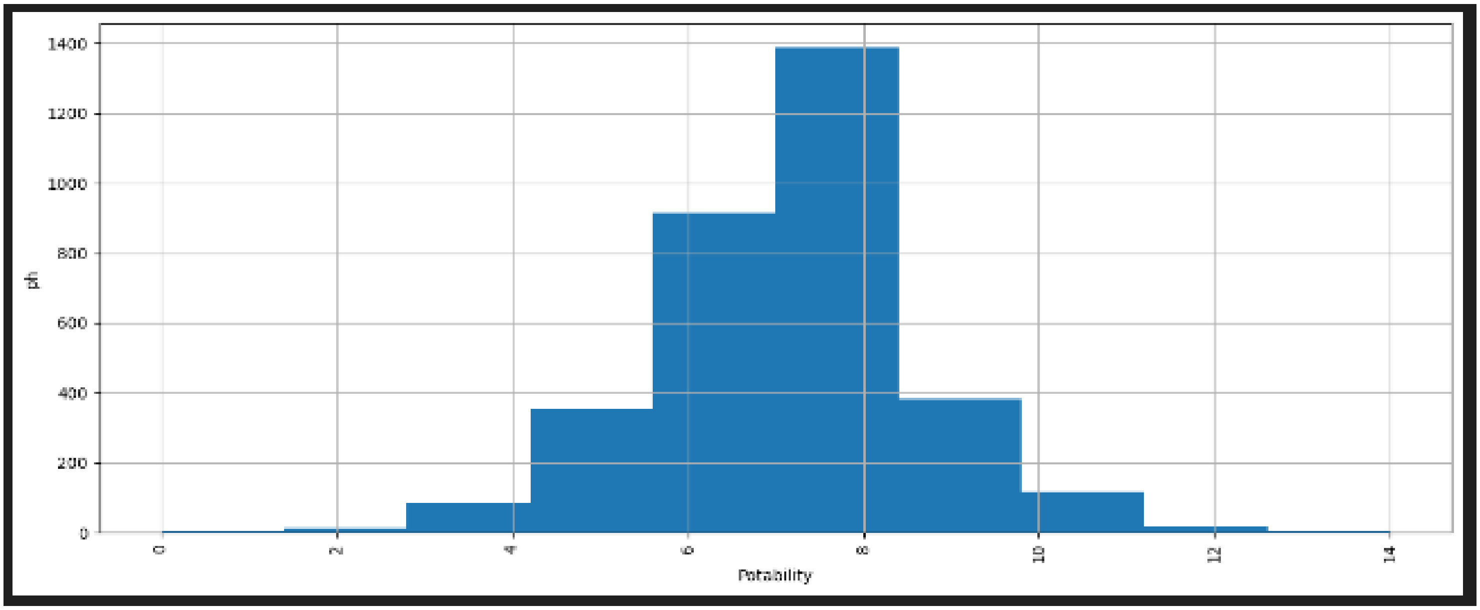


FIG 6.1.3 SCREENSHOT ON POTABILITY GRAPH

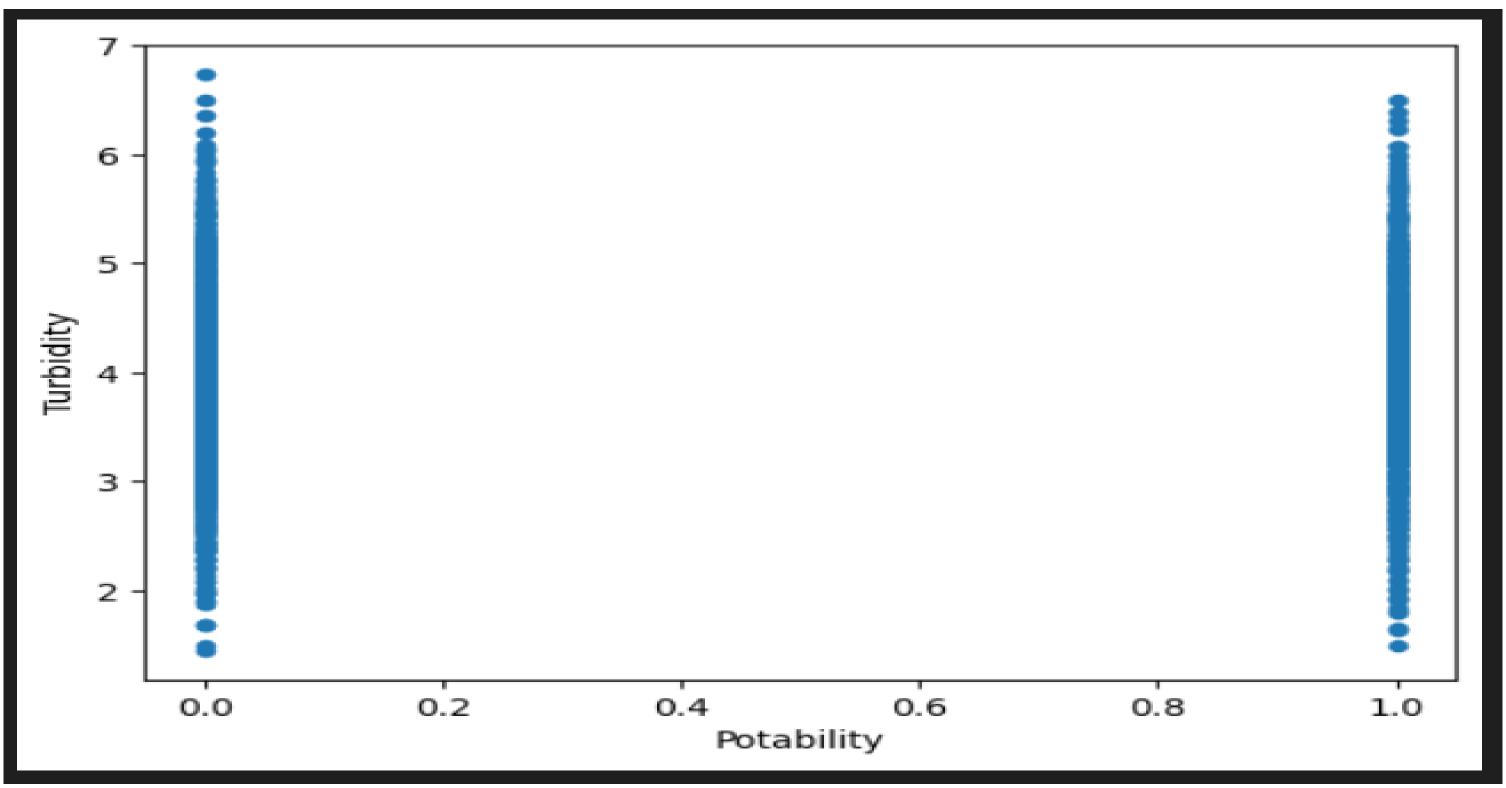


FIG 6.1.4 SCREENSHOT OF TURBIDITY GRAPH

6.2 RESULT & DISCUSSION

A comprehensive river surveillance system is indispensable for assessing the condition of river ecosystems and ensuring their long-term health and sustainability. This system operates through persistent data collection and analysis of various parameters crucial for understanding water quality and ecosystem dynamics. Key parameters include water temperature, pH levels, dissolved oxygen concentrations, turbidity, and concentrations of pollutants such as heavy metals, nutrients, and organic compounds.

The primary function of the river surveillance system is to enable the identification of pollution sources and trends in water quality over time. By continuously monitoring these parameters along different points in the river, the system can pinpoint areas of concern, detect changes in water quality patterns, and track the effectiveness of remediation efforts. Moreover, the system facilitates the early detection of environmental threats such as chemical spills or algal blooms, enabling prompt response measures to mitigate their impact and prevent further contamination. Stakeholders, including government agencies, environmental organizations, water management authorities, and local communities, rely on the data collected by the river surveillance system to make informed decisions about managing water resources. By analyzing monitoring data, stakeholders can identify areas where pollution control measures need to be intensified, optimize the allocation of water resources to different uses, and implement strategies to minimize the adverse effects of human activities on river ecosystems.

Furthermore, a well-executed river water monitoring system serves to safeguard and sustainably manage riverbed ecosystems. By ensuring the continuous monitoring of water quality parameters, the system helps maintain the ecological health of rivers, preserving biodiversity, and ecosystem functions. Additionally, by providing reliable data on water quality, the system supports efforts to ensure the safety of water supplies for human consumption, protecting public health and promoting environmental justice.In essence, the purpose of a robust river surveillance system goes beyond mere data collection; it is about safeguarding the health of river ecosystems, promoting sustainable water management practices, and ensuring the availability of clean and safe water resources for both ecological and human needs now and in the future.

CHAPTER 7
CONCLUSION AND FUTURE WORK

7.1 CONCLUSION

Potability determines the quality of water, which is one of the most important resources for existence. Traditionally, testing water quality required an expensive and time-consuming lab analysis. This study looked into an alternative machine learning method for predicting water quality using only a few simple water quality criteria. To estimate, aset of representative supervised machine learning algorithms was used. It would detect water of bad quality before it was released for consumption and notify the appropriate authorities It will hopefully reduce the number of individuals who drink low-quality water, lowering the risk of diseases like typhoid and diarrhea. In this case, using a prescriptive analysis based on projected values would result in future capabilities to assistdecision and policy makers.

7.1 FUTURE ENHANCEMENTS

As technology advances and our understanding of water quality dynamics improves, the future of predicting water quality parameters using machine Learning holds great promisefor more accurate, timely, and actionable insights. Continuous research, interdisciplinary collaboration, and technological innovation will play key roles in shaping the future of water quality prediction.

7.1.1 Integration of Advanced Sensors

Incorporating advanced sensors and Internet of Things (IoT) devices can enhance data collection accuracy and frequency. These sensors may include hyperspectral sensors, nanotechnology-based sensors, or even autonomous underwater vehicles for more comprehensive monitoring.

7.1.2 Real-time Monitoring and Early Warning Systems

Future systems could focus on real-time monitoring with the development of more sophisticated early warning systems. This involves implementing machine learning models that can rapidly detect anomalies or predict potential water quality issues before they escalate.

APPENDIX

A1.SOURCE CODE

```
import pandas as pd import
numpy as np import seaborn
as sns
import matplotlib.pyplot as plt import
warnings
warnings.filterwarnings("ignore")
from sklearn.preprocessing import LabelEncoderfrom
sklearn import metrics
df = pd.read_csv('water_potability.csv')
df.head(10)
df.shape
df.isnull().sum()
df['ph']=df['ph'].fillna(df['ph'].mean())
df['Sulfate']=df['Sulfate'].fillna(df['Sulfate'].mean())
df['Trihalomethanes']=df['Trihalomethanes'].fillna(df['Trihalomethanes'].mean()) plt.figure(figsize=(15,15))
sns.heatmap(df.corr(), annot=True)
plt.show()
plt.figure(figsize=(15, 6))
plt.xticks(rotation=90)
df.ph.hist()
plt.xlabel('Potability')
plt.ylabel('ph')
plt.plot()
sns.pairplot(data=df)
plt.figure(figsize=(15, 6))
```

```
plt.xticks(rotation=90)
df.ph.hist()
plt.xlabel('Potability')
plt.ylabel('ph') plt.plot()

df.plot(kind="scatter", x="Potability", y="Turbidity")X =
df.drop(['Potability'], axis=1)
Y = df['Potability'] import
tensorflow as tf
from sklearn.model_selection import train_test_splitfrom
sklearn.preprocessing import StandardScaler
# Assuming X and Y are your features and target variable# Split the
data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, Y, test_size=0.2, random_state=42)

# Standardize the features
scaler = StandardScaler()
X_train_scaled = scaler.fit_transform(X_train)X_test_scaled =
scaler.transform(X_test)
# Define the architecture of the ANN modelmodel
= tf.keras.Sequential([
    tf.keras.layers.Dense(64, activation='relu', input_shape=(X_train_scaled.shape[1],)),tf.keras.layers.Dense(32,
    activation='relu'),
    tf.keras.layers.Dense(1, activation='sigmoid')
])
# Compile the model
model.compile(optimizer='adam', loss='binary_crossentropy', metrics=['accuracy'])
```

```
# Train the model
history = model.fit(X_train_scaled, y_train, epochs=300, batch_size=32,validation_split=0.2)
# Evaluate the model on the test set
# test_loss, test_acc = model.evaluate(X_test_scaled, y_test)#
print('Test Accuracy:', test_acc)
from tensorflow.keras.models import load_model# Load
the model from the .h5 file
loaded_model = load_model("final_model.h5") import numpy as np#
Assuming 'loaded_model' is the loaded TensorFlow model
# User input data#
user_input =
np.array([[8.316766,214.373394,22018.417441,8.059332,356.886136,363.266516,18.4
36524,100.341674,4.628771]])
user_input = np.array([[3.716080,129.422921,18630.057858,6.635246,333.775777,592.885359,15.1
80013,56.329076,4.500656]])
#user_input = np.array([[8.099124,224.236259,19909.541732,9.275884,333.775777,418.606213,16.8
68637,66.420093,3.055934]])
# Perform any necessary preprocessing (e.g., scaling) on the user input data
# Here, you can use the same scaler object you used for scaling the training data user_input_scaled =
scaler.fit_transform(user_input)  # Assuming 'scaler' is the scalerobject used for scaling

# Make predictions using the loaded model
predictions = loaded_model.predict(user_input_scaled)
#rounded_predictions = [1 if predictions >= 0.5 else 0 for predictions in predictions]
```

```
# Print the predictions
print(predictions)
if predictions > 0.5:
    print("1")
else:
    print("0")

import numpy as np

# Assuming 'loaded_model' is the loaded TensorFlow model# User
input data
# user_input = np.array([[5.126763,230.603758,11983.869376,6.303357,333.775777,402.883113,11.1
68946,77.488213,4.708658]])
# user_input = np.array([[7.874671,195.102299,17404.177061,7.509306,333.775777,327.459760,16.1
40368,78.698446,2.309149]])
user_input = np.array([[9.419510,175.762646,33155.578218,7.350233,333.775777,432.044783,11.0
39070,69.845400,3.298875]])
# Perform any necessary preprocessing (e.g., scaling) on the user input data
# Here, you can use the same scaler object you used for scaling the training data user_input_scaled =
scaler.transform(user_input) # Assuming 'scaler' is the scalerobject used for scaling
# Make predictions using the loaded model
predictions = loaded_model.predict(user_input_scaled)
#rounded_predictions = [1 if predictions >= 0.5 else 0 for predictions in predictions]# Print the
predictions
```

```
print(predictions)
if predictions > 0.5:
    print("1")
else:
    print("0")
from sklearn.model_selection import train_test_split
X_train,X_test,Y_train,Y_test=train_test_split(X,Y,test_size=0.2,random_state=70)
print(X_train.shape,X_test.shape,Y_train.shape,Y_test.shape)
from sklearn.linear_model import LinearRegression
lr_model = LinearRegression() lr_model.fit(X_train,
Y_train)
from sklearn.linear_model import LogisticRegressionlr=
LogisticRegression(random_state=0) lr.fit(X_train, Y_train)
from sklearn.tree import DecisionTreeRegressor
regressor = DecisionTreeRegressor()
regressor.fit(X_train, Y_train)
from sklearn.ensemble import RandomForestRegressorreg =
RandomForestRegressor(n_estimators=100,
                                 random_state=0)
reg.fit(X_train, Y_train)
from sklearn.naive_bayes import GaussianNB
classifier = GaussianNB() classifier.fit(X_train,
Y_train)
if int(Result1[0])==1:
    print('Output : ', 'Potable ')
else:
    print('Output : ', ' Not potable')
```

import pickle

Dump the trained Naive Bayes classifier with PickleDT_pkl_filename

= 'DecisionTree2.pkl'

Open the file to save as pkl file DT_Model_pkl =

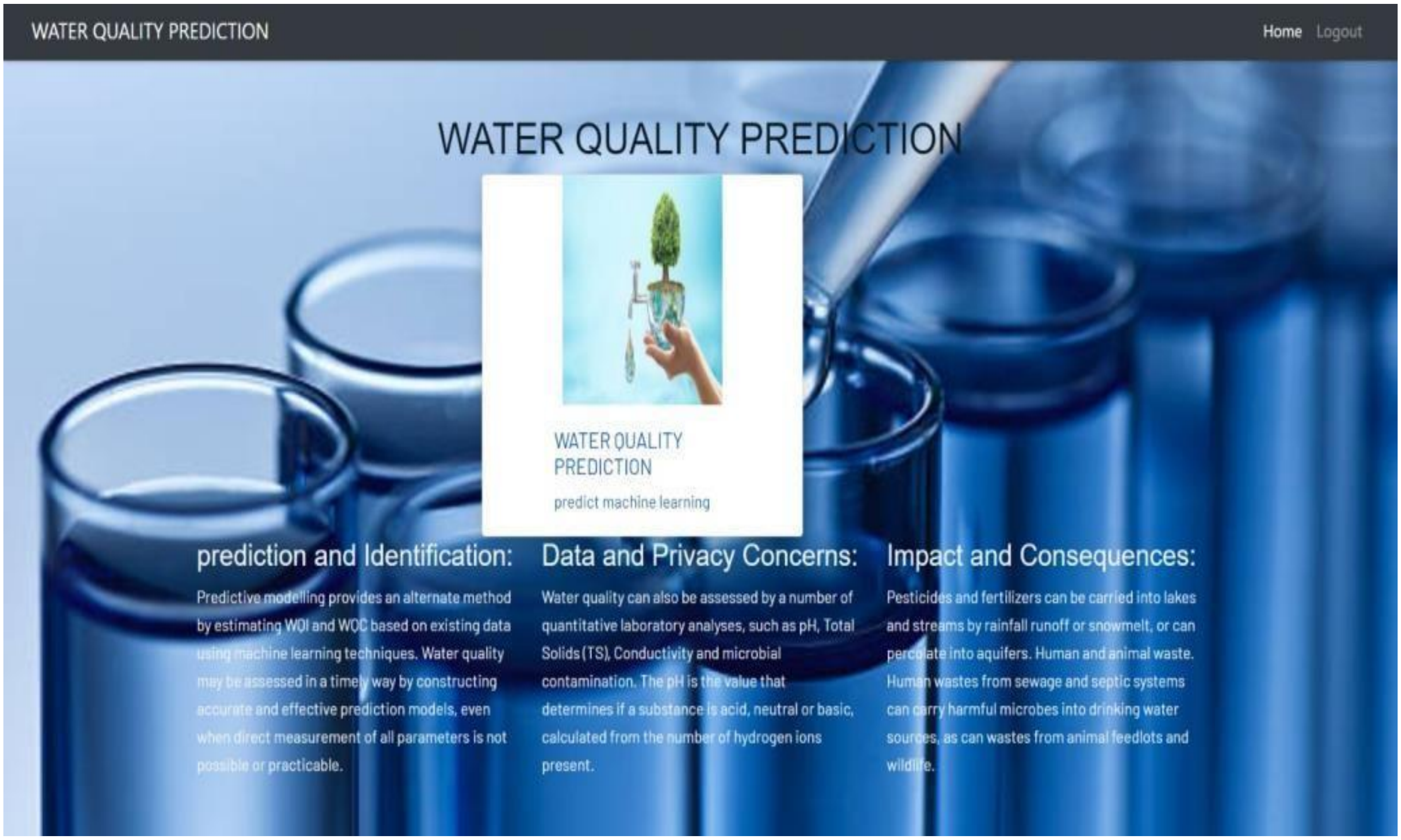
open(DT_pkl_filename, 'wb')pickle.dump(regressor,

DT_Model_pkl)

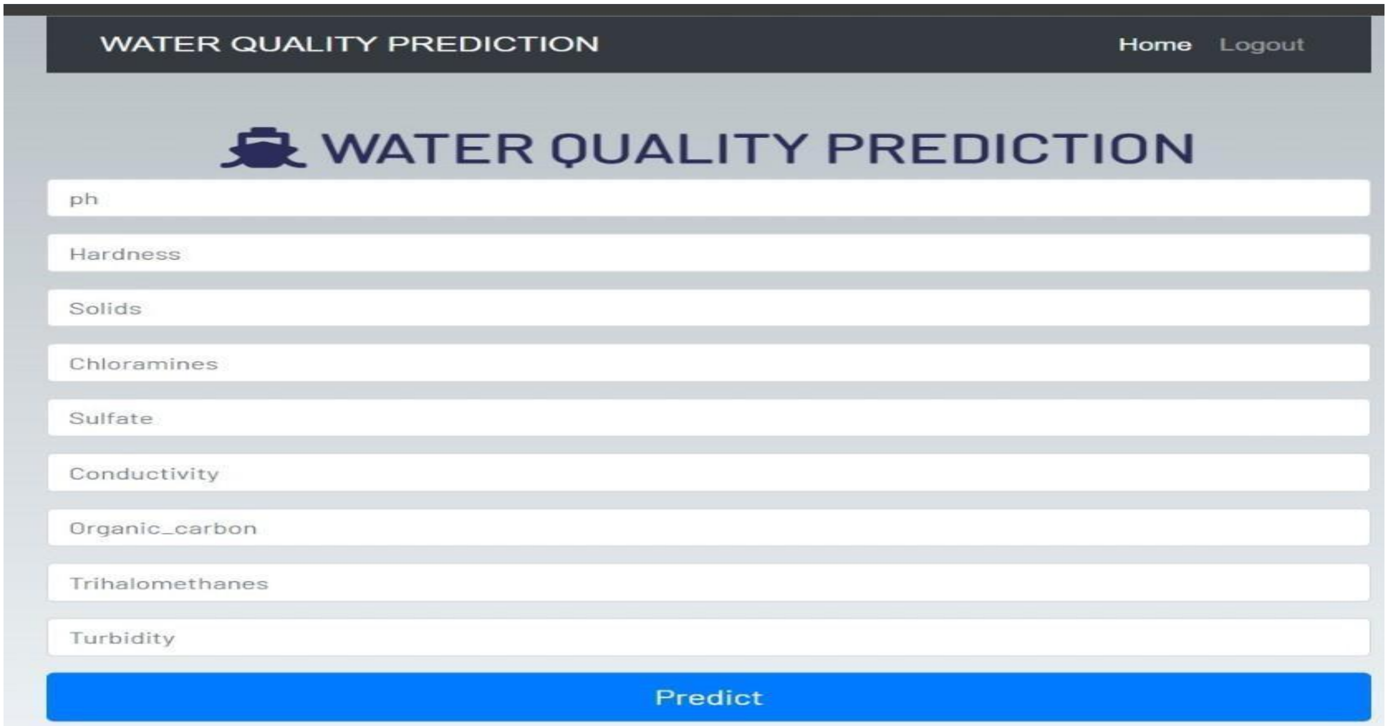
Close the pickle instances

DT_Model_pkl.close()

A2.SCREENSHOTS



A.2.1 SCREENSHOT OF START PAGE



A.2.2 SCREENSHOT OF LOGIN PAGE

A.2.3 SCREENSHOT OF RESULT PAGE

A3.PLAGIARISM REPORT

Authenticated by ANTIPLA plagiarism checker
Date of issuance 2024-03-21 14:25:52
Accessed via on www.antip.la

Result

This document contains 11% plagiarism. This means that the author has copied data from public sources when writing this work.

Analysis

Result	**11%**
Document title	RIVER WATER QUALITY MONITORING SYSTEM
Content hash	09d874cbb7b05f2d930a949c7c4b9fab
Date	2024-03-21 14:24:16
Check time	18 seconds
Character count	10,000
Special character count	55
Word count	1,420
Number of plagiarized words	352

Plagiarism sources

https://newsletter.x-mol.com/paper/1371979198058938368?adv
https://research.vit.ac.in/publication/crop-yield-prediction-usi...
https://newsletter.x-mol.com/paper/1699175900400865280
https://ojs.uma.ac.id/index.php/jite/article/view/2599
https://pubmed.ncbi.nlm.nih.gov/35145205/
https://escholarship.org/content/qt9zs4n73w/qt9zs4n73w.pdf?t=q0n...
https://www.linkedin.com/pulse/delivering-accuracy-progress-blue...

REFERENCE

[1] PCRWR. National Water Quality Monitoring Programme, Fifth Monitoring Report(2005– 2006); Pakistan Council of Research in Water Resources Islamabad: Islamabad,Pakistan,2007.

[2] Kangabam, R.D.; Bhoominathan, S.D.; Kanagaraj, S.; Govindaraju, M. Developmentof a water quality index (WQI) for the Loktak Lake in India. Appl. Water Sci. 2017, 7, 2907–2918. [CrossRef]

[3] Thukral, A.; Bhardwaj, R.; Kaur, R. Water quality indices. Sat 2005, 1, 99.

[4] Srivastava, G.; Kumar, P. Water quality index with missing parameters. Int. J. Res. Eng. Technol. 2013, 2, 609–614.

[5] The Environmental and Protection Agency, "Parameters of water quality," Environ. Prot., p. 133, 2001.

[6]Y. Chen, L. Song, Y. Liu, L. Yang, and D. Li, ''A review of the artificial neural network models for water quality prediction,'' Appl. Sci., vol. 10, no. 17, p. 5776, Aug.2020.

[7] T. Rajaee, S. Khani, and M. Ravansalar, ''Artificial intelligence-based single and hybrid models for prediction of water quality in rivers: A review,'' Chemometric Intell. Lab. Syst., vol. 200, May 2020, Art. no. 103978.

[8] S. Wechmongkhonkon, N. Poomtong and S. Areerachakul, "Application of Artificial Neural Network to Classification Surface water Quality", World Academy of science, Engineering and Technology, vol. 6, (2012) September 29.

[9] D. Sheppard, T. D Tsegaye, W. Tadesse, D. Mckay and T. L Coleman, "The Application of Remote Sensing, Geographic Information Systems, and Global positioning System Technology to improve Water Quality in Northern Alabama", Centre for hydrology, soil climatology, and Remote sensing Department of Plant and soil science, (2001).

[10] X. Yunrong and L. Jiang, "Water quality optimization", Second international workshop on prediction using LS-SVM with particle swarm knowledge discovery and Data mining, IEEE, (2009).

[11] C. Zhou, G. Liang, G. Haibing and P. Chuanyong, "Pattern classification and prediction of water quality by neural network with particleswarm optimization", proceedings of the 6th world congress on intelligent control and automation, IEEE, (2006) June 21-23, Dalian, China.

[12] R. Barzegar, J. Adamowski, and A. A. Moghaddam, ''Application of wavelet- artificial intelligence hybrid models for water quality prediction: A case study in Aji- Chay River, Iran,'' Stochastic Environ. Res. Risk Assessment, vol. 30, no. 7, pp. 1797– 1819, Oct. 2016.

[13] M. Ehteram, S. Q. Salih, and Z. M. Yaseen, ''Efficiency evaluation of reverse osmosis desalination plant using hybridized multilayer perceptron with particle swarmoptimization,'' Environ. Sci. Pollut. Res., vol. 27, no. 13, pp. 15278–15291, May 2020.

[14] Z. Yaseen, M. Ehteram, A. Sharafati, S. Shahid, N. Al-Ansari, and A. El-Shafie, ''The integration of nature-inspired algorithms with least square support vector regression models: Application to modeling river dissolved oxygen concentration,'' Water, vol. 10, no. 9, p. 1124, Aug. 2018.

[15] M. Bayatvarkeshi, M. A. Imteaz, O. Kisi, M. Zarei, and Z. M. Yaseen, ''Applicationof M5 model tree optimized with excel solver platform for water quality parameter

[16] Nikhil Kumar Koditala ; Purnendu Shekar Pandey,‖ Water Quality Monitoring System Using IoT and Machine Learning‖, IEEE paper, August 2018.

[17] M. Zennaro, A. Floros, G. Dogan, T. Sun, Z. Cao, C. Huang, M. Bahader and A. Bagula, "On the design of a water quality wirelesssensor network (WQWSN): an application to water quality monitoring in Malawi", international conference on parallel processing workshops, DOI 10.1109/ICPPW.2009.57.

[18] S. Peng and X. Zhao, "A review of uncertainty methods employed in water quality modeling", International Conference on Environmental Science and Information Application Technology, DOI 10.1109/ESIAT.2009.227.

[19] C. Zhu and Z. Hao, "Fuzzy Neural Network Model and its Application in water Quality Evaluation", International conference on environmental science and information application technology. DOI 10.1109/ESIAT.2009.45.

[20] M. Makhtar, D. Neagu and M. J. Ridley, "Binary Classification Models Comparison: on the Similarity of Datasets and Confusion Matrix for Predictive Toxicology Applications", 2nd International Conference on Information Technology in Bio and Medical Informatics ITBAM2011, (2011), pp. 108. 14. J. B. T. Ismail, "Assessment of Kinta River Water Quality by using Multivariate Technique", (2011).