# ARTIFICIAL INTELLIGENCE BASED DATA-DRIVEN DECISION SUPPORT SYSTEM IN WASTEWATER TREATMENT PLANTS

A PROJECT REPORT

submitted by

**MANNA ANN MARIYAM (MGP18CS087)**
**SHERIN JOHN (MGP18CS116)**
**SIAN GIJO (MGP18CS118)**
**WINNY ANNA VARKEY (MGP18CS130)**

to

the APJ Abdul Kalam Technological University in partial fulfillment of the requirements for the award of the Degree

of

Bachelor of Technology
in
*Computer Science and Engineering*

SAINTGITS
LEARN.GROW.EXCEL

**Department of Computer Science and Engineering Saintgits College of Engineering, Pathamuttom**

JUNE 2022

# DECLARATION

We, undersigned hereby declare that the project report **“Artificial Intelligence Based Data-Driven Decision Support System in Wastewater Treatment Plants”**, submitted for partial fulfillment of the requirements for the award of degree of Bachelor of Technology of the APJ Abdul Kalam Technological University, Kerala is a bonafide work done by us under the supervision of **Er. Nisha Joseph**. This submission represents our ideas in our own words and where ideas or words of others have been included, we have adequately and accurately cited and referenced the original sources. We also declare that we have adhered to ethics of academic honesty and integrity and have not misrepresented or fabricated any data or idea or fact or source in our submission. We understand that any violation of the above will be a cause for disciplinary action by the institute and/or the University and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been obtained. This report has not been previously formed the basis for the award of any degree, diploma or similar title of any other University.

Place: **Kottayam**

Date: **03-06-2022**

Names & Signature of the students:

(1) **Manna Ann Mariyam**

(2) **Sherin John**

(3) **Sian Gijo**

(4) **Winny Anna Varkey**

**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**
**SAINTGITS COLLEGE OF ENGINEERING, PATHAMUTTOM**

## CERTIFICATE

This is to certify that the report entitled **Artificial Intelligence Based Data-Driven Decision Support System in Wastewater Treatment Plants** submitted by **Manna Ann Mariyam, Sherin John, Sian Gijo, Winny Anna Varkey** to the APJ Abdul Kalam Technological University in partial fulfillment of the requirements for the award of the Degree of Bachelor of Technology in Computer Science and Engineering is a bonafide record of the project work carried out by him/her under my guidance and supervision. This report in any form has not been submitted to any other University or Institute for any purpose.

Internal Supervisor

**Er. Nisha Joseph**

Project In-Charges

1. **Er. Nisha Joseph**
2. **Dr. Reni K Cherian**

HEAD OF THE DEPARTMENT

**Dr. Anju Pratap**

# CONTENTS

Contents Page No.

# LIST OF TABLES

# LIST OF FIGURES

# ACKNOWLEDGEMENT

First and foremost, we express our gratitude to our principal, **Dr. Josephkunju Paul C,** Principal, Saintgits College of Engineering for providing us with excellent ambiance that laid a potentially strong foundation for our work. We would like to convey our profound gratitude to **Dr. Anju Pratap**, Head of the Department of Computer Science and Engineering, Saintgits College of Engineering, for being a consistent source of support and strength throughout our project. We express our sincere gratitude to the project coordinators, **Dr. Reni K Cherian** and **Er. Nisha Joseph** of the Computer Science and Engineering Department provides us with all the facilities, valuable and timely suggestions, and constant supervision for the successful completion of our project. We also thank our Supervisor **Er. Nisha Joseph**, Assistant Professor, Computer Science and Engineering Department for his supervision, encouragement, help and support. We thank our college management for making all necessary arrangements throughout the project work. We are highly indebted to all the faculties of the department for their valuable guidance and instant help and for being with us. We extend our heartfelt thanks to our parents, friends and well-wishers for their support and timely help. Last but not the least, we thank Almighty God for helping us throughout the project.

# ABSTRACT

The Wastewater Treatment Plant (WWTP) plays a crucial role in our society. It involves a combination of complex processes used to treat and remove pollutants from water. All the decisions in WWTP are conventionally taken by skilled and qualified plant operators with the necessary training and education in order to get the job done right. There can be a considerable amount of error that can occur when critical decisions are taken by these operators. In order to tackle this and to improve efficiency and accuracy, a Decision Support System (DSS) can be used as traditional methods of decision making by human operators are considerably less efficient. Water quality parameters such as pH, Hardness, Solids, Chloramines, Sulfate, Conductivity, Organic Carbon, Trihalomethanes, Turbidity are used to determine the purity status of water being considered. Best results from WWTPs can be resolved by observing these parameters and controlling them. These nine parameters serve as features in the dataset used. The proposed system is a Machine Learning based DSS built on Decision Tree Algorithm that will predict the potability of water using historical data, which will aid the plant operators in making daily operational decisions at the WTTPs. It will reduce a lot of errors that can be caused and is thus very efficient and cost effective. The system offers a user-friendly and simple User Interface which can be operated with limited technological knowledge. Daily water quality parameter readings can be given as input to the system which would generate a binary output indicating the potability. Since this is a system that purely works on the basis of historical data given training time, it can be used in a variety of industries with slight modifications made.

# CHAPTER 1
# INTRODUCTION

Changes in water supply and wastewater problems put India's population at risk. Climate change-related effects on the monsoon are having, and will continue to have, major ramifications for agriculture, making India even more vulnerable. The broad geographical layout of India allows it to reclaim a significant volume of wastewater. Almost 80% of the water supply is returned to the ecosystem as effluent. If not managed properly, this can pose a serious environmental and health risk, but proper management could assist water managers in meeting the city's water demands. Furthermore, when it comes to wastewater treatment systems that are designed to treat wastewater, it can be noted that these wastewater treatment plants are faced with their own issues. Some of the problems faced are those related to energy consumption due to the highly complex nature of wastewater treatment, environmental footprint and labor shortages. To successfully treat wastewater, you'll need efficient and productive personnel who are knowledgeable about the industry. Unfortunately, there aren't enough employees to go around. This leads to the need for automation in the system that fills in the gaps of labor shortages as well as lower the energy consumption.

## 1.1 General Background

Clean water is one of the most essential natural resources on the earth because it is necessary for life. The elimination of contaminants from wastewater before it reaches aquifers or natural bodies of water such as rivers, lakes, and seas is known as wastewater treatment. The type and concentration of pollutants detected in the water, as well as its intended usage, determine any distinction between clean and contaminated water. Water is said to be polluted in broad terms when it contains enough contaminants to make it inappropriate for a specific use, such as drinking, irrigation, or fishing. Although natural factors have an impact on water quality, the term pollution usually refers to contamination caused by human activity. Water pollution is mostly caused by the discharge of contaminated wastewater into bodies of water, and wastewater treatment is a key component of water pollution prevention. Global drought map reveals that many parts of the world lack adequate water. Water treatment plants (WWTP) are built to accelerate the natural purification process. The natural process is overburdened by billions of humans and significantly more wastewater. The volume of wastewater produced without wastewater treatment would create havoc, as it does today in developing countries. Over 80% of all wastewater is discharged without treatment around the world. In the countries that do have water treatment facilities, they use a variety of technologies to purify water and return it to the environment, with the goal of keeping humans and the environment safe and thriving. Water demand is predicted to quadruple by 2035 in the water sector; as a result, huge amounts of additional energy will be used for water supply and wastewater treatment unless plant energy efficiency is improved. All communities, particularly those in water-scarce locations, must ensure that they have effective water treatment mechanisms in place so that treated water can be reused or recycled.

## 1.2 Relevance

Wastewater, which is basically used water, is also an important resource, especially in locations where there are frequent droughts and water shortages. Wastewater treatment is essential for two reasons: restoring water sources and protecting the environment from toxins. In wastewater, contaminants from both home and corporate use can be present. Chemical compounds and pathogens in wastewater, if left untreated, can impair the health of animals, plants, and birds living in or near water. It can also contaminate crops and drinking water, putting people's quality of life at stake. Wastewater treatment is critical for the health of a variety of ecosystems. When wastewater is adequately treated, it can be used as a supply of water for a variety of uses. When wastewater is treated properly, the maximum quantity of water can be utilized rather than being wasted.

Traditionally, all decisions in a WWTP are made by experts and technicians, making it vulnerable to a variety of errors. Beginner mistakes with wastewater flow, pumping, inlet and outlet hydraulics, detention time, aeration, and sludge treatment reveal a clear lack of professional exposure. Lack of expertise manifests itself in the frequent copying of foreign solutions and old initiatives, without regard for local circumstances or other significant impacts. In this field, a lack of systematic specialist education might have serious consequences. If wrong decisions are taken at various stages, this will directly affect the treatment of water. When this ill treatment is released into the environment, it will badly affect the ecosystem and its inhabitants.

So the requirement of an Artificial Intelligence based Decision Support System that will assist plant operators to make decisions effectively is necessary.

## 1.3 Socio Economic Importance

For decisions to be taken effectively in WWTP, requires highly skilled and educated personnel. Hiring such individuals can be very expensive. In order to help these technicians, highly precise machinery is also required. Instead, if a single intelligent system is established, it can do the precise decision making and thus reduce the cost that might be needed for correction of the wrong decisions taken by humans. This cost efficient mechanism can contribute to the economic well being of the nation.

Water contamination is becoming a major hindrance to any society's long-term sustainable growth. Considering India’s capability in treating wastewater, this potential presents the importance of replenishing water resources that are highly in demand. Hence, establishing effective water treatment methodologies can benefit the society by providing treated water that contributes to the general well-being of human health and ecosystem.

# CHAPTER 2

# LITERATURE SURVEY

Inorder to perform wastewater in any country, it is vital to abide by the nation's water quality standards. Table 2.1 , shows the water quality standards that are most commonly used in India [1].

| **Designated-Best-Use** | **Class of water** | **Criteria** |
|---|---|---|
| Drinking Water Source without conventional treatment but after disinfection | A | ● Total Coliforms Organism MPN/100ml shall be 50 or less<br>● pH between 6.5 and 8.5<br>● Dissolved Oxygen 6 mg/l or more<br>● Biochemical Oxygen Demand 5 days 20ºC celsius 2 mg/l or less |
| Outdoor bathing (Organised) | B | ● Total Coliforms Organism MPN/100ml shall by 500 or less pH between 6.5 and 8.5 DO 5 mg/l or more<br>● Biochemical Oxygen Demand 5 days 20ºC 3 mg/l or less |
| Drinking water source after conventional treatment and disinfection | C | ● Total Coliforms Organism MPN/100ml shall be 5000 or less pH between 6 to 9 DO 4 mg/l or more<br>● Biochemical Oxygen Demand 5 days 20ºC 3 mg/l or less |
| Propagation of wildlife and fisheries | D | ● pH between 6.5 to 8.5 DO 4 mg/l or more<br>● Free Ammonia (as N) 1.2 mg/l or less |
| Irrigation, Industrial Cooling, Controlled Waste disposal | E | ● pH between 6.0 to 8.5<br>● Electrical Conductivity at 25ºC micro mhos/cm Max. 2250<br>● Sodium absorption Ratio Max. 23<br>● Boron Max. 2 mg/l |
| | Below-E | Not meeting A, B, C, D & E criteria |

**Table 2.1 Criteria for wastewater treatment**

Techniques for wastewater treatment in this sector includes separation of floatable solids, flow and load equalization, sedimentation, biological treatment for reduction of BOD, effluent chlorination, dewatering and disposal of residuals.

Article [2] presents a review of control techniques that have been applied to wastewater treatment processes in the past four decades. The 21st century is characterized by an evolving control approach in process industries. Control technology has become a top focus in the industry as a result of stringent environmental and health laws, as well as a requirement for cost-effective wastewater treatment methods.

a. Linear Control - LC theory deals with systems that can be formalized either in numerical or continuous form.

b. Linearizing Control - Nonlinearities have long been the origin of control upset issues. The use of linearization techniques have solved this problem. There are two fundamental types of linearizing control:

- Input-output feedback linearization , where the output relationship to the input system of the feedback-loop system must be linear, and
- Input-state feedback linearization, where the states of the closed-loop states and the system input are linear.

c. Nonlinear Control - Many wastewater treatment processes exhibit non-linear behavior. Classically, their control used a linear model obtained by linearization around an operating point. But incase of highly nonlinear discontinuous processes, conventional linear control may prove to be inefficient. Thus nonlinear geometric theory with gain scheduling and multivariable control which studies different types of models for both their nonlinearity, their robustness and their statistical efficiency , can begin to address the issue.

d. Artificial Intelligence- Based Control Methods - Recent times have witnessed a large scale use of artificial intelligence- based methods for large scale industrial processes. Several AI techniques,such as fuzzy logic, neural networks, and genetic algorithms, could address uncertainty in environmental models and have been applied. In several WWTPs, there are also many strategies for control that are a combination of AI approaches with combination of conventional linear methods or linearizing control techniques.

In WWTP [3], a fuzzy logic based control system was introduced for diagnosis and control of an anaerobic WWTP. Three variables ( methan reproduction, hydrogen concentration in the gas phase and total alkalinity ratio) were selected as the most adequate for organic overload identification.

In [4], different machine learning techniques were used to model a soft-sensor to predict weather conditions for advanced control of Wastewater Treatment Plants such as Support Vector Machine, k-nearest neighbors, Decision Trees, Random Forest and Gaussian Naive Bayes. The system obtained approximately 85% accuracy in the weather soft-sensor with two machine learning algorithms: KNN and Random Forests.

Article [5] proposes a Supervised Committee Fuzzy Logic (SCFL) model that adopts an artificial neural network as a non linear assembler to integrate three FL models of Takagi-Sugeno, Mamdani and Larsen such that individual FL model outputs are input to ANN to derive new outputs. The result is that non linear combinations of FL models under supervision showed better performance than their individual linear combination in predicting water quality parameters.

In [6], an ANN with a feed-forward, back propagation learning paradigm was adopted to predict the effluent water quality of Habesha brewery WTP. ANN was found to predict the same with correlation coefficient (R) between observed and predicted output values reaching 0.969.

In paper [7], the performance of Nicosia WWTP in terms of effluent biological oxygen demand (BOD), chemical oxygen demand (COD) and total nitrogen (TN) was modeled by AI models of Feed Forward neural network (FFNN), Adaptive neuro fuzzy inference system (ANFIS), Support Vector Machine (SVM) and a classical multi-linear regression (MLR). Simple averaging (SAE), weighted (WAE) and neural network ensemble (NNE) techniques were further employed to enhance prediction performances of individual models. ANFIS showed to have better performance than other single models. In the case of ensemble predictions, SAE, WAE and NNE increased the AI modeling efficiency by 14%, 20% and 24% respectively making NNE more robust and efficient.

Work [8] uses an artificial neural network (ANN) black-box modeling approach to acquire the knowledge base of a real wastewater plant and then use it as a process model. The study signifies that the ANNs was capable of capturing the plant operation characteristics with a good degree of accuracy.

Paper [9], proposes a knowledge-data-driven flexible switching controller(KDFSC) designed and analyzed to achieve reliable control performance. The result demonstrates that the proposed KDFSC can achieve excellent control performance.

In paper [10], a Fuzzy partial least square-based dynamic Bayesian network (FPLS-DBN) is proposed to improve the mobility in WWTPs. A fuzzy partial least squares (FPLS) is introduced by using a fuzzy system to extract nonlinear features from process data. In addition, a dynamic extension is included by embedding augmented matrices into Bayesian networks to fit the uncertainty and time- varying characteristics. The study shows the superiority of FPLS-DBN in modeling performance for an actual industrial WWTP application.

Paper [11] compares three different data driven methods, extreme learning machine (ELM), least square support vector machine (LS-SVM) regression method and back propagation neural network (BPNN) for prediction of outlet TP concentration. The result shows that the ELM method has remarkably superior performance on TP out prediction than peer models.

# CHAPTER 3
# OBJECTIVES AND PROPOSED INNOVATION

## 3.1 Objectives

The main objectives of this project are to:

I. Understand and analyze various approaches and mechanisms involved in wastewater treatment .

II. Identify and understand the application of artificial intelligence in the predictive analysis of WWTPs.

III. Study and find differences in various approaches and existing systems related to intelligent wastewater treatment systems.

IV. Build an Intelligent Decision Support System for WWTPs using Artificial Intelligence by employing a robust and accurate model

V. Create a User Interface for the interaction of the plant operator with the system to generate the desired application.

## 3.2 Proposed Innovation

The proposed system consists of an Intelligent Decision Support System in WWTPs , that can aid in the daily operational decision making in the plant. This system makes use of Machine Learning and various learning algorithms to achieve this intelligence. The model employed will provide superior performance in robustness and accuracy than conventional learning models by using ensemble data intelligence models which is a combination of several base models.

Literature Survey clearly shows that Machine Learning and other related systems are far more intelligent and efficient in taking decisions than the conventional methods. The use of this system can also reduce the energy consumption and the cost tremendously.

# CHAPTER 4
# HYPOTHESIS, DESIGN AND METHODOLOGY

## 4.1 Hypothesis

Artificial Intelligence based Decision Support System for Wastewater Treatment Plants will outline a model that learns historical data collected from the plants and use this knowledge in labeling new and unseen data. This will aid the plant operators in decision making at the treatment plants by predicting the potability of water based on the corresponding water quality parameters that are given as input to the system. The model is built using the Decision tree algorithm.

## 4.2 Design

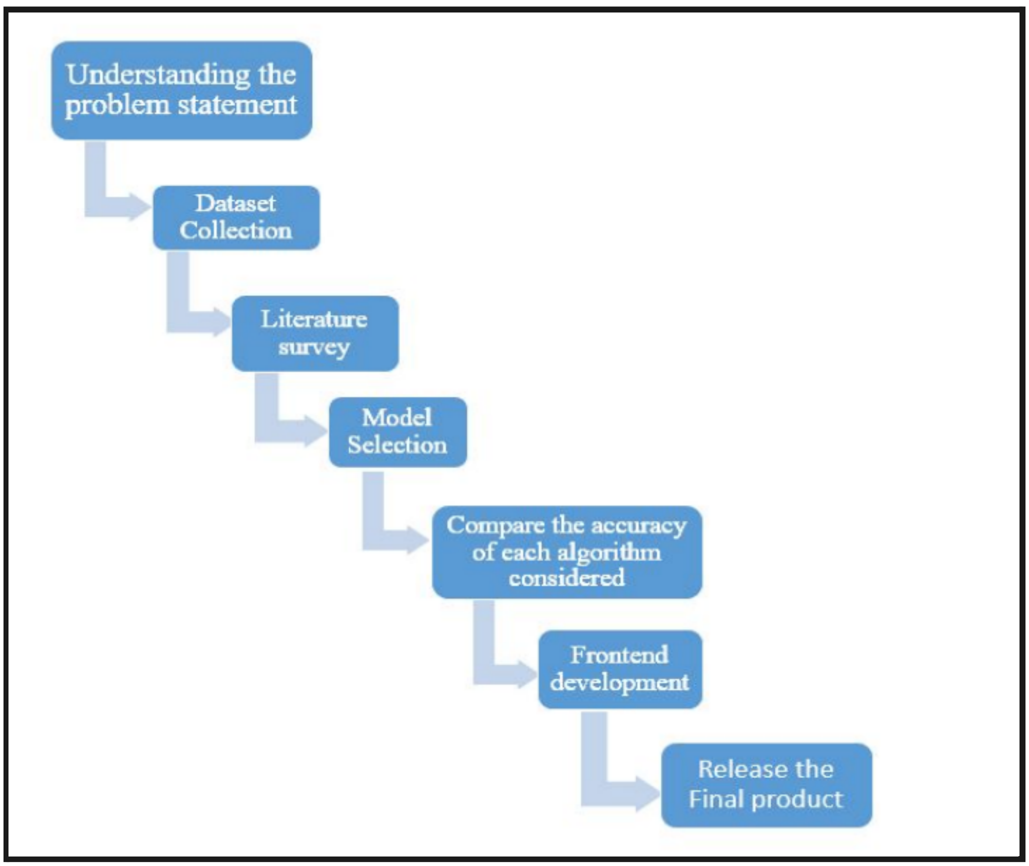


**Figure 4.1 Project Design Flowchart**

The overall project design is shown in Figure 4.1.

Understanding the problem statement

Clearly understanding the problem statement is very essential for correct formulation of the result. The main focus was on understanding the key areas of concern faced in WWTPs and using current technologies to solve them effectively. During this study the problems statements that were evident were:

1. Conventional methods used in mainstream WWTPs are very time consuming, bear a lot of cost and are not efficient.
2. Other similar systems already in use algorithms that have low accuracy.

Dataset Collection

A dataset that had measures of various features of wastewater was required. The main features that were needed were pH, Hardness, Solids, Chloramines, Sulfate, Conductivity, Organic Carbon, Trihalomethanes, Turbidity and other such features whose levels in water are critical factors that determine its purity. Since dataset from WWTPs were difficult to be sourced due to certain technical and privacy policy issues , our study mainly used open datasets. The dataset used is [12].

Literature Survey

The first stage of the project mainly focused on knowing more about the various processes and requirements of WWTPs and closely analyzing the decision making systems used. This study was able to observe a wide range of differing technologies that are under use. These ranged from decisions taken by technicians, conventional methods such a use of linear control, nonlinear control theory, linearizing control etc. As the complexities of the processes increased and as the conventional methods became less effective, the search began for other techniques that can ease the process of decision making. Thus artificial intelligence and machine learning were popularized in this field. From then, many studies and researches were conducted by various researchers and scholars to explore more into these domains and to use them in different processes inside WWTPs.

Model Selection, Compare the accuracy of each algorithm considered

Inorder to create the required Intelligent Decision Support System (IDSS) there are various machine learning algorithms that are available for use. Each algorithm, its uses and performance measures are to be clearly studied. A list of suitable algorithms is to be created and then trained with training dataset and then use validation dataset and calculate the accuracy of each. The algorithm that gives the maximum accuracy and also the one that best suits the problem statement will be selected.

Frontend Development

The user interface will be a web platform that can provide features like to input the user data, display results of predictions and also visualize various results.

Release the Final Product

After development of all the components of the product, it will be released into the industry for use.

### 4.3 Methodology

A decision support system (DSS) aids in the making of decisions, judgements, and courses of action in a business or firm. A data science system sifts through and analyses large amounts of data, generating complete data that can be utilized to solve problems and make decisions. Machine-learning approaches can help with wastewater treatment process design by speeding up process simulation, which is an important step in both constructing new plants and refining the design of existing ones. The proposed system utilizes this functionality by modeling an Artificial Intelligence model built on Decision tree algorithm. The model considers nine features namely pH, Hardness, Solids, Chloramines, Sulfate, Conductivity, Organic Carbon, Trihalomethanes and Turbidity which are water quality parameters that determine the status of purity level of the water being considered. A plant operator can make use of the simple User Interface to enter daily water quality parameter readings. The model will then predict this unseen new data as potable water or not which will be indicated as a binary output.

The dataset used was initially preprocessed to handle the null values in the dataset. This was done by replacing the null values by the mean of all the remaining values of that attribute. This preprocessing was done for all the features. Then the dataset was split into testing and validation sets as 90% for training and 10% for validation. The model was developed using Jupyter notebook by importing libraries such as Numpy, pandas, scikit learn and so on. Scikit learn is an efficient machine learning library in python that offers a selection of tools for modeling including classification, regression, etc.

The initial study for this project compared the accuracy for various algorithms such as Random Forest, Artificial Neural Network, Support Vector Machine and so on.

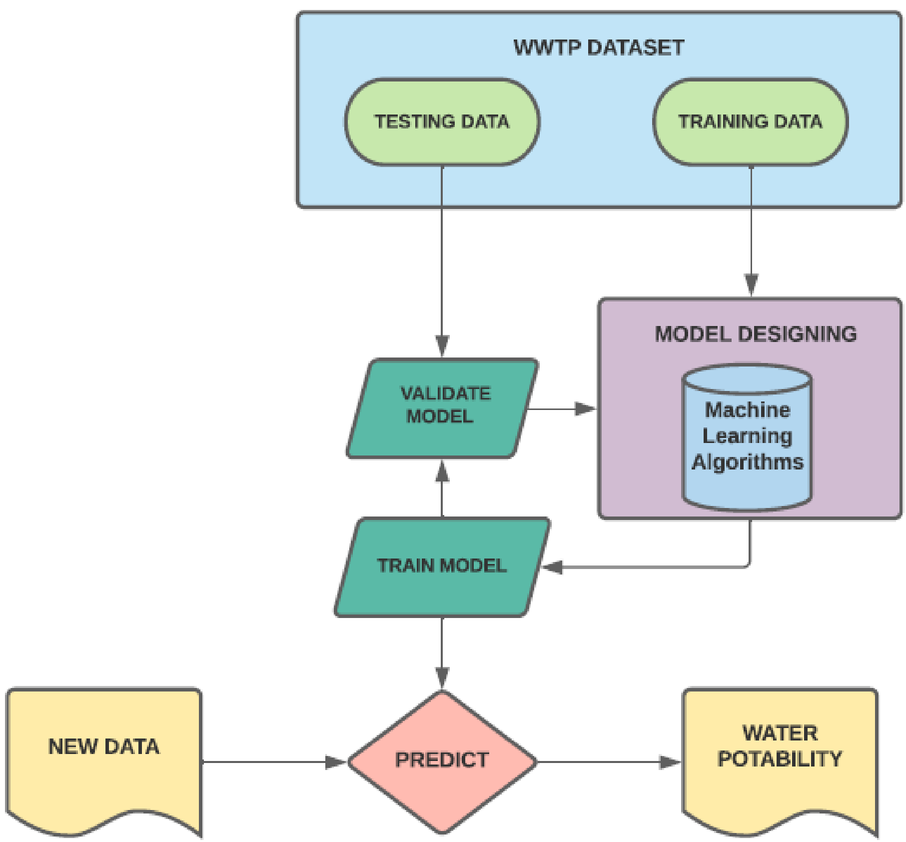


**Figure 4.2 Schematic Representation of Prediction model**

#### 4.3.1 Dataset

The dataset used for this project is [12]. It consists of nine features that control the water quality determination. They are namely-

1) pH

   pH is a logarithmic scale that indicates the acidity or alkalinity of a solution. Where 7 is neutral, low values are more acidic, and high values are more alkaline.

2) Hardness

   Hard water is water with a high mineral content. Hard water occurs when water infiltrates through limestone, choke, or gypsum deposits that are primarily composed of calcium carbonate and magnesium, bicarbonate, and sulfate. Hard drinking water has moderate health benefits.

3) Solids

   It is a measure of the total dissolved content of all inorganic and organic substances present in a liquid in a suspended form of molecules, ionizations, or fine particles.

4) Chloramines

   Chloramines are disinfectants used to treat drinking water. Chloramines are most commonly formed when ammonia is added to chlorine to treat drinking water.

5) Sulfate

   Sulfates are naturally occurring anions in both surface and groundwater freshwater. Sulfate levels above 250 mg / l can add bitterness to water.

6) Conductivity

   Conductivity is the result of the presence of mineral salts of elements such as sodium, calcium and magnesium. When dissolved in water, these salts produce free ions that can conduct electricity in the water.

7) Organic Carbon

   Total organic carbon is a measure of the amount of organic compounds present in a water sample. The higher the carbon or organic content, the more oxygen is consumed. High organic matter content means increased growth of microorganisms that contribute to the depletion of oxygen supply.

8) Trihalomethanes

   Total trihalomethanes are a group of disinfection by-products formed when chlorine compounds used to disinfect water react with other naturally occurring chemicals in water.

9) Turbidity

   Turbidity is a measure of the relative transparency of a liquid. This is the optical property of water, a measure of the amount of light scattered by substances in the water when it is exposed to light through a water sample.

### 4.3.2 Decision Tree Algorithm

Decision Trees are a kind of Supervised Learning where the learning takes the form of a tree-structure consisting of decision nodes and leaves. The root node is also a decision node, but at the topmost level. The data gets continuously split according to certain parameters. The decisions or final outcomes are represented by the leaves. And the data is separated at the decision nodes. We start from the root of the tree when using Decision Trees to predict a class label for a record. The values of the root attribute and the record's attribute are compared. We follow the branch that corresponds to that value and jump to the next node based on the comparison. The examples are classified using decision trees by sorting them down the tree from the root to a leaf/terminal node, with the leaf/terminal node providing the classification.

Each node in the tree represents a test case for some feature, with each edge descending from the node corresponding to the test case's possible solutions. This is a cyclical procedure that occurs for each subtree rooted at the new node.

The outcomes can be categorical (for example, Yes or No) or continuous-valued. Numerous algorithms can be used to construct Decision Trees namely ID3 algorithm, CART algorithm and so on.

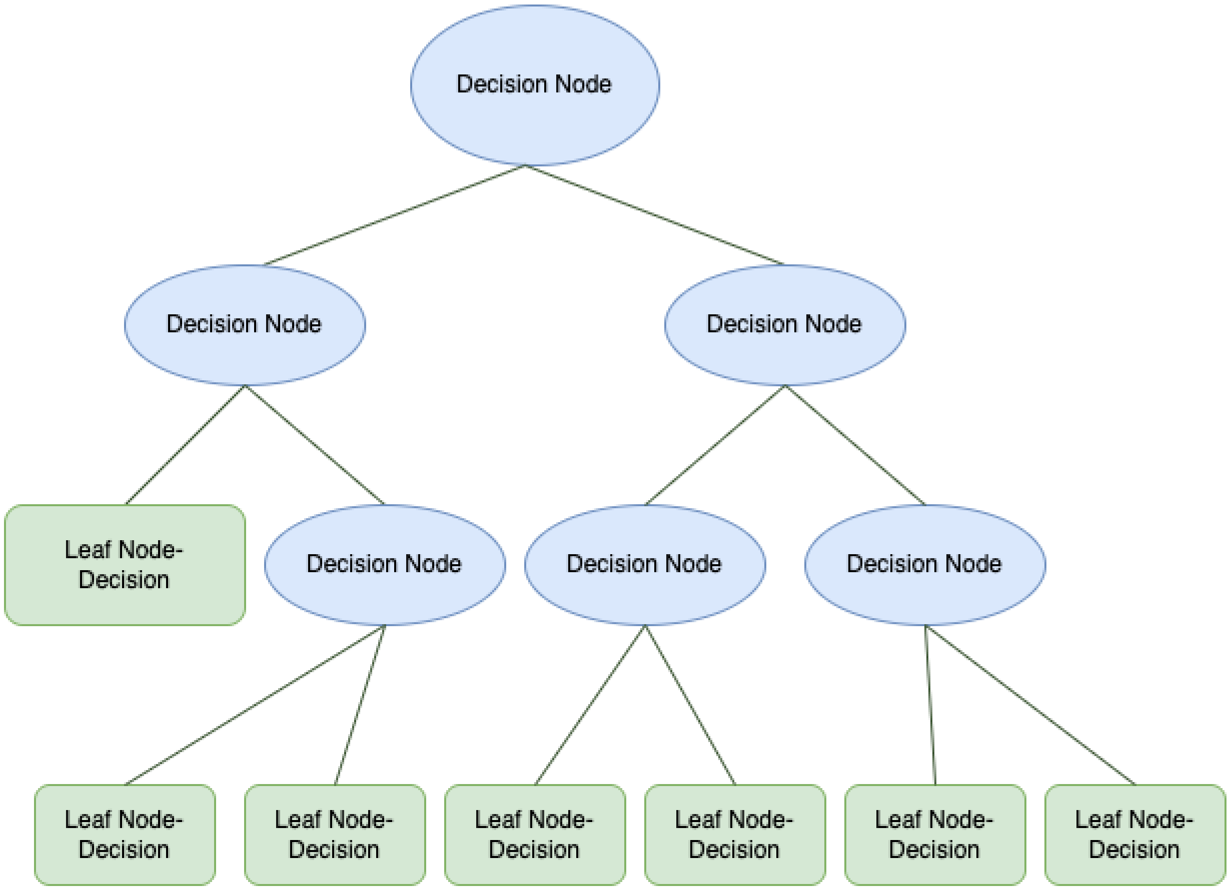


**Figure 4.3 Decision Tree Structure**

### 4.3.3 User Interface

The system provides a simple and user-friendly interface that can be used by plant operators with minimal technical knowledge. The user interface was created using Streamlit by Python language which is an open source web application framework for creating apps for machine learning applications. There are nine input text boxes where the user can enter water quality parameter readings such as pH, Solids, Turbidity, etc. and a submit button 'Classify' to submit the input test data to the model. The model then predicts the potability by displaying a binary 0/1 output where a binary 1 indicates that the water being tested is potable and a 0 indicating otherwise.

# CHAPTER 5
# RESULTS AND DISCUSSIONS

The model provides sufficiently accurate results. The study has compared various Machine Learning Algorithms to find their accuracy. The algorithms considered were K-Nearest Neighbor , Linear SVC, Logistic Regression, Support Vector Classifier, Gaussian Naive Bayes, Decision Tree, Random Forest and Stochastic Gradient Descent. The accuracy obtained for each of the above mentioned algorithms are listed in Figure 5.1.

| | MODEL | SCORE |
|---|---|---|
| 1 | Support Vector Machine | 0.594512 |
| 2 | KNN | 0.649390 |
| 3 | Logistic Regression | 0.583841 |
| 4 | Random Forest | 0.615854 |
| 5 | Linear SVC | 0.535061 |
| 6 | Naive Bayes | 0.571646 |
| 7 | SGD | 0.637195 |

**Table 5.1 Comparison of accuracy of several Machine learning algorithms**

Among all the algorithms compared, it is Decision Tree Algorithm , that gave the maximum accuracy of about 72%.

```
In [68]: from sklearn.tree import DecisionTreeClassifier
         dt = DecisionTreeClassifier(criterion = 'entropy', random_state = 42)
         dt.fit(X_train, y_train)
         dt_pred_train = dt.predict(X_train)

In [69]: # Evaluation on Training set
         dt_pred_train = dt.predict(X_train)

         print('Training Set Evaluation F1-Score=>',f1_score(y_train,dt_pred_train))

         Training Set Evaluation F1-Score=> 1.0

In [70]: # #Evaluating on Test set
         dt_pred_test = dt.predict(X_test)
         acc_dec_tree=accuracy_score(y_true=y_test , y_pred=dt_pred_test)
         acc_dec_tree
          #print('Testing Set Evaluation F1-Score=>',f1_score(y_test,dt_pred_test))

Out[70]: 0.7134146341463414
```

**Figure 5.2 Decision Tree model accuracy**

## 5.1 Model Output

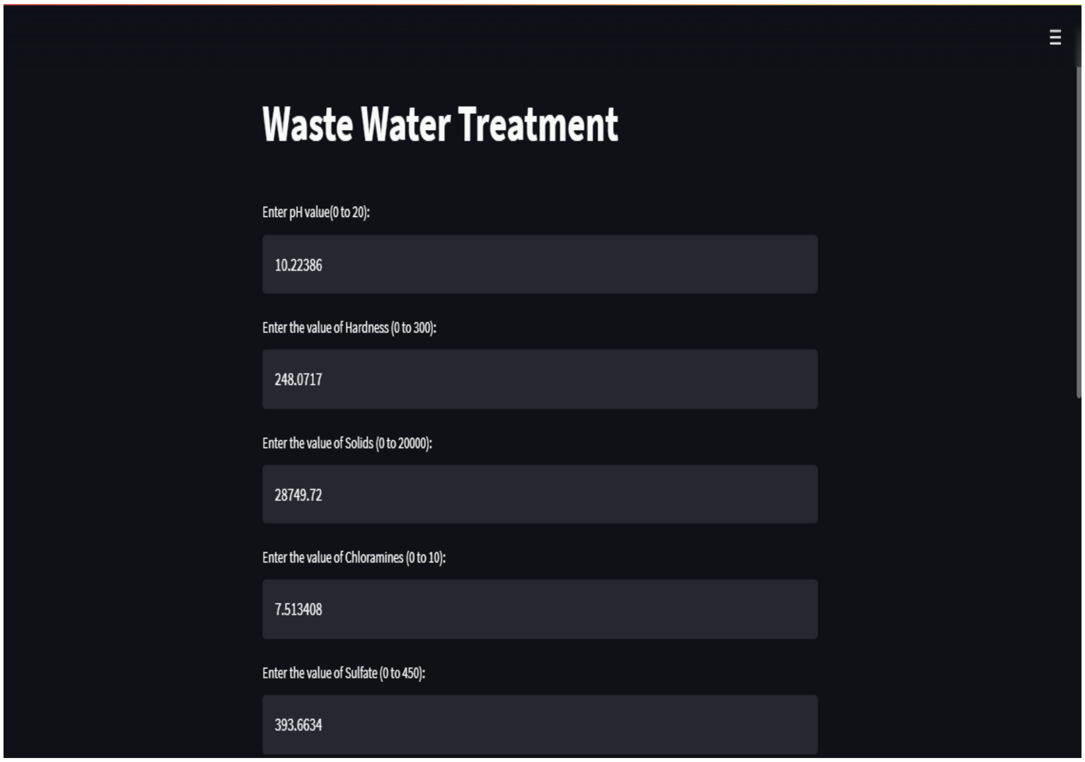


**Figure 5.3  User Interface: Testing example 1 (a)**

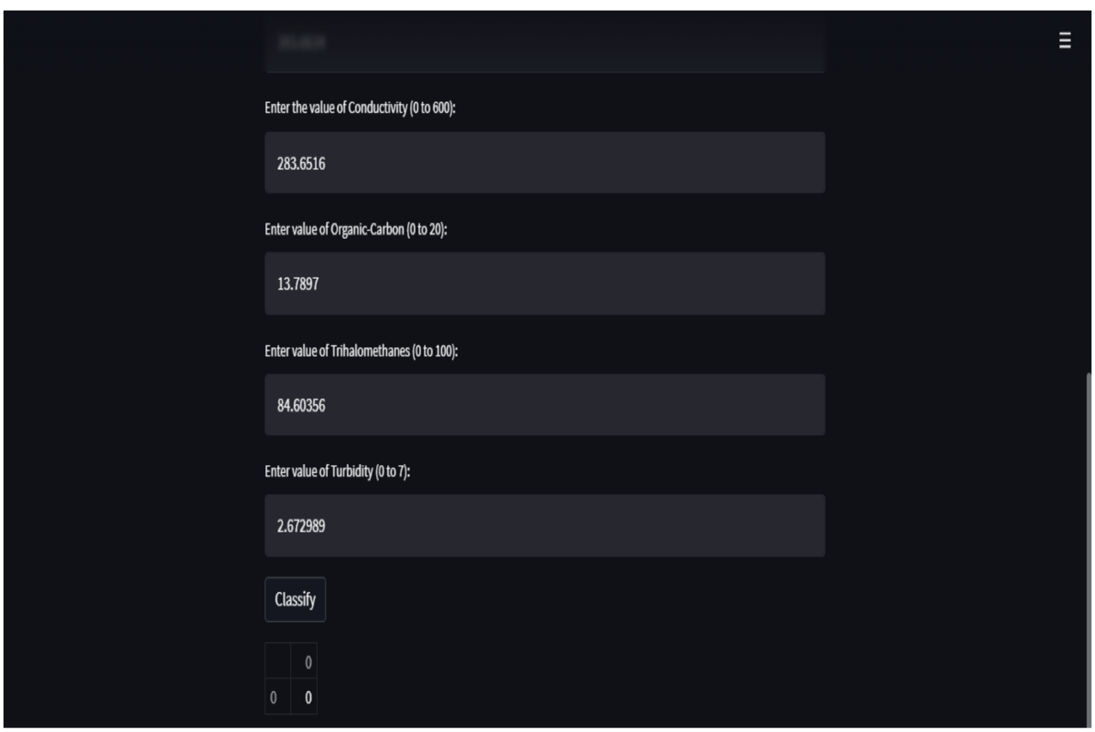


**Figure 5.4 User Interface: Testing Example 1 (b)**

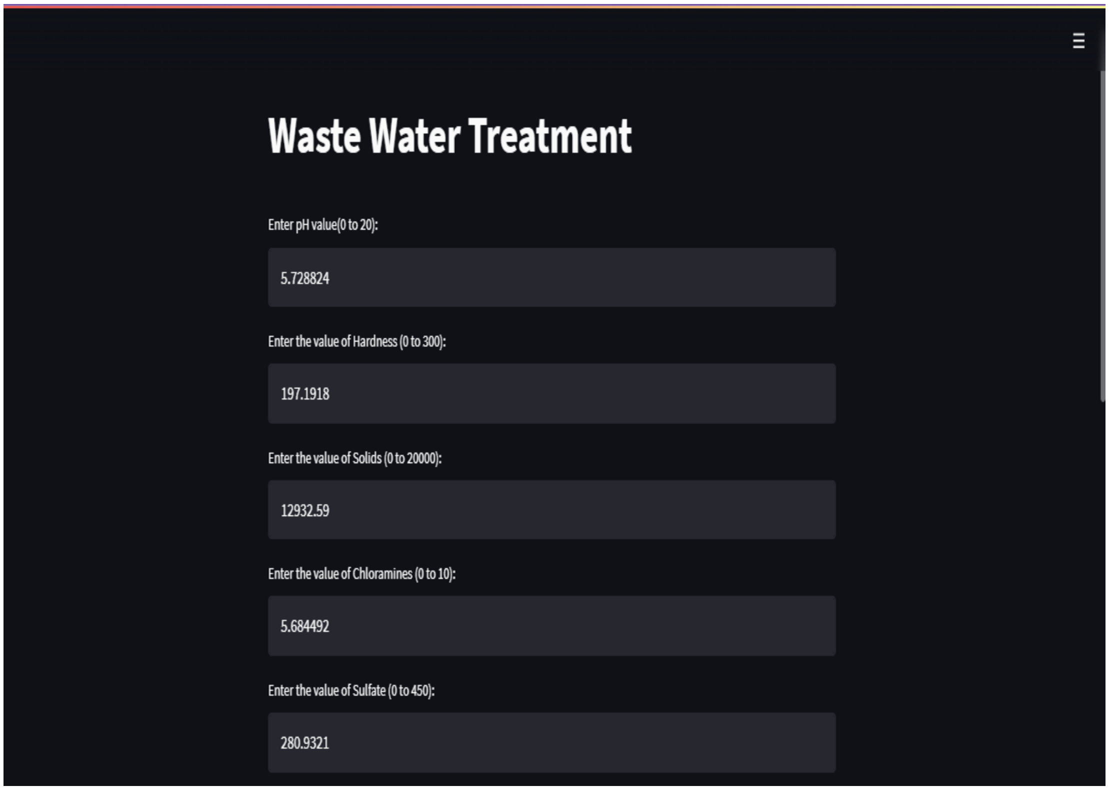


**Figure 5.5 User Interface: Testing Example 2 (a)**

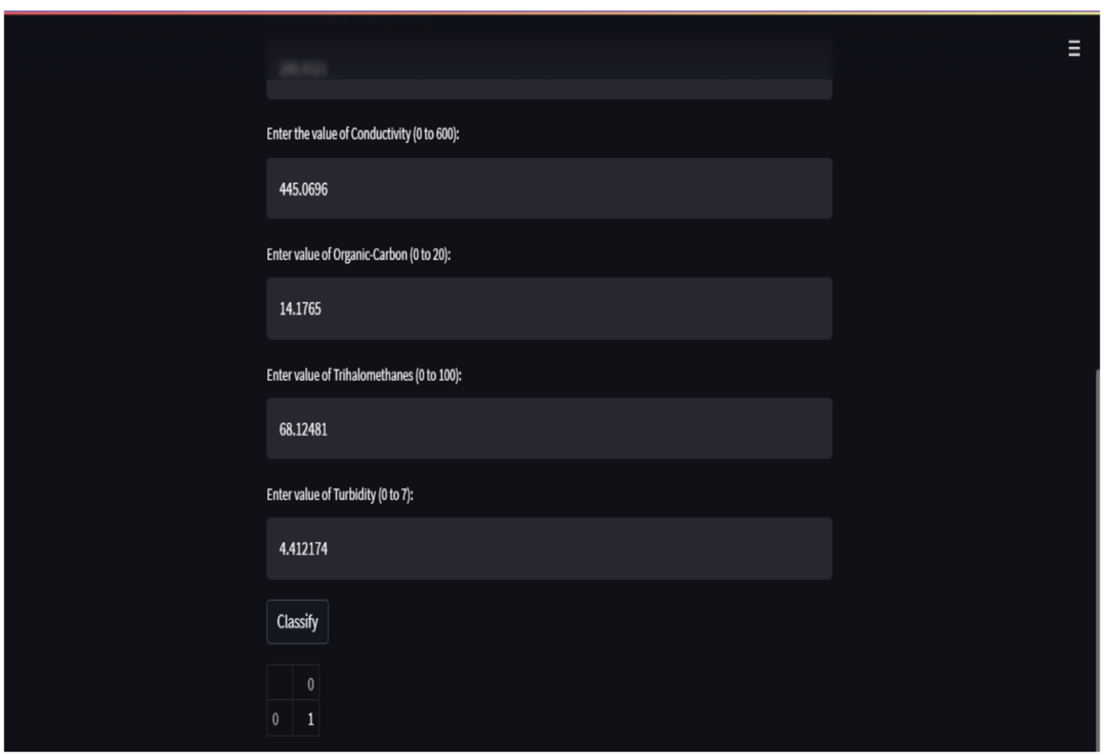


**Figure 5.6 User Interface: Testing Example 2 (b)**

The decision tree created by the model is represented in figure 5.5 with maximum depth of 3.

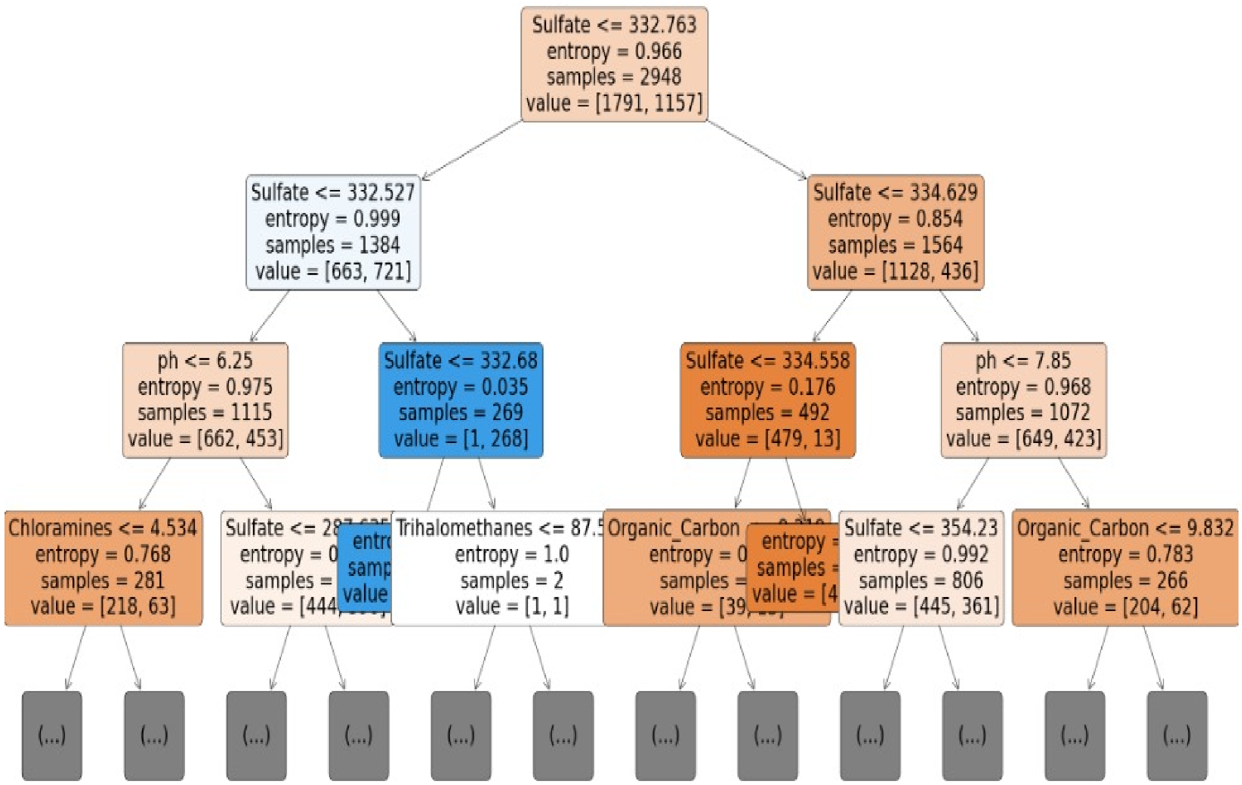


**Figure 5.7 Decision Tree Visualization**

It can be seen from the decision tree visualization (Figure 5.5) that the model has chosen the feature Sulfate as the best splitting algorithm and hence forms the root decision node. Subsequent tree growth can be studied from the figure.

### 5.2 Model Performance Evaluation

The model's performance was evaluated using accuracy metrics such as F1 score and Accuracy score. From figure 5.2, it can be seen that the model has about 72% accuracy on the test data.

When it comes to handling classification problems, the confusion matrix is a prominent tool. A confusion matrix is a table that is frequently used to describe a classification model's performance on a set of test data for which the true values are known. Confusion matrices are used to depict counts based on actual and predicted values. The output "TN" stands for True Negative, and it displays the number of correctly identified negative cases. Similarly, "TP" stands for True Positive, which denotes the number of correctly identified positive examples. The terms "FP" and "FN" stand for False Positive and False Negative values, respectively. "FP" stands for False Positive value, which is the number of actual negative examples classified as positive, and "FN" stands for False Negative value, which is the number of actual positive examples classified as negative [13].
The confusion matrix of the project's model is shown in figure 5.8.

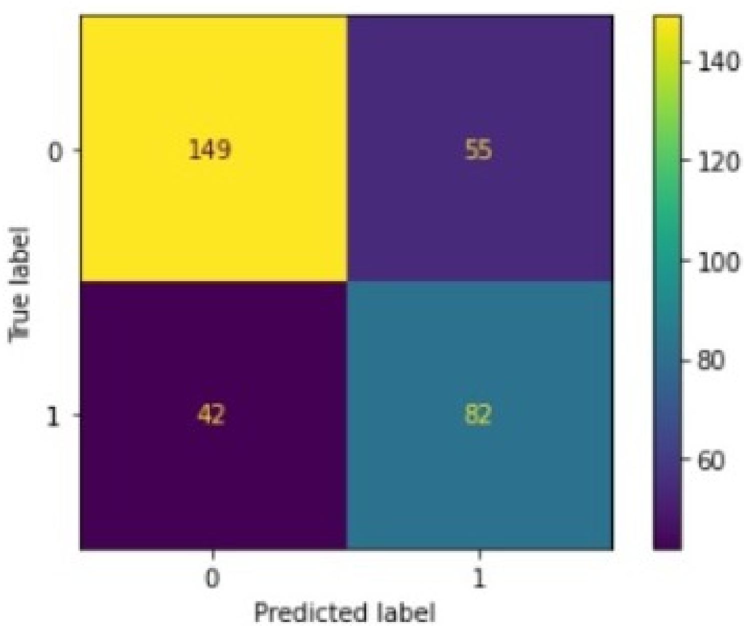


**Figure 5.8 Confusion Matrix of the model**

# CHAPTER 6
# FORMULATION OF WORK PLAN

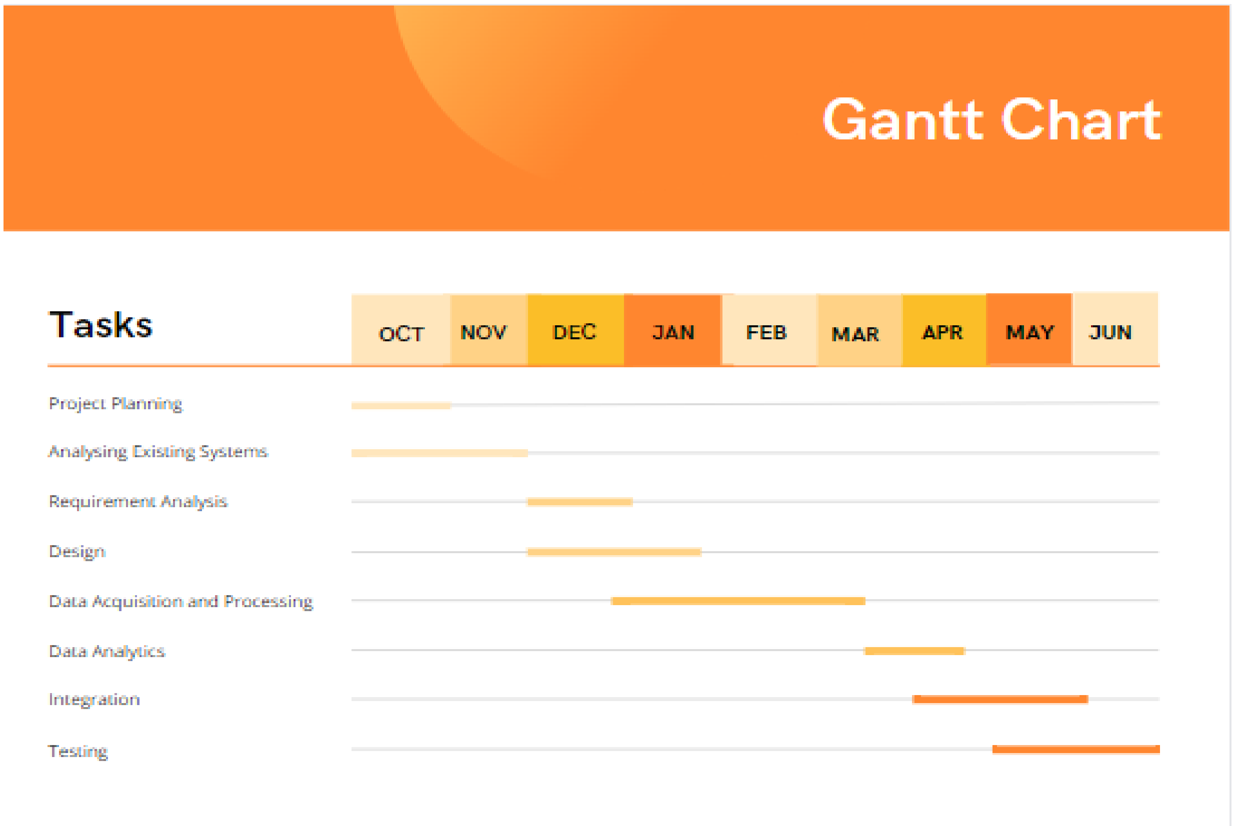


**Figure 5.1 Gantt Chart**

Figure 5.1 demonstrates the overall work plan of our project. The project is mainly divided into 3 phases namely Preliminary Phase (25%), Design Stage (50%) and Completion Phase (25%). Preliminary phase has been completed along with zeroth and first phase reviews. The final phase mainly includes final paperwork and release of the final product.

# CHAPTER 7
# CONCLUSION AND FUTURE SCOPE

Water is one of the essential resources on earth as it is necessary for life. Wastewater is used water that might contain a lot of contaminants. Since wastewater contains hazardous elements, it is required to be treated before it is released into the environment or reused for other purposes. Thus wastewater treatment proves to be critical.

Earlier decisions in WWTPs solely depended on the knowledge and expertise of the experts and technicians in charge and thus were error prone. Thus the need for an intelligent system was very evident.

Even though conventional methods proved to be useful in small scale facilities, as the processes become more complex and exhibit nonlinear nature, these methods proved to be insufficient.

As the literature study demonstrates, no single model is superior to others in all circumstances, and the performance of several models varies depending on the state of each WWTP. The research conducted has provided a lot of insight into intelligent wastewater treatment and various approaches to it which will be beneficial to carry on future works in this domain. The knowledge obtained will be utilized in modeling an efficient and dynamic Decision Support System that would prove to be superior in terms of robustness and accuracy for handling the non linear complexities of WWTPs.

# REFERENCES


1. “Vikaspedia Domains.” Accessed January 13, 2022. https://vikaspedia.in/energy/policy-support/environment-1/water/water-quality-and-standards.
2. Iratni, Abdelhamid, and Ni-Bin Chang. “Advances in Control Technologies for Wastewater Treatment Processes: Status, Challenges, and Perspectives.” IEEE/CAA Journal of Automatica Sinica 6, no. 2 (March 2019): 337–63. https://doi.org/10.1109/JAS.2019.1911372.
   Accessed on 31 December 2021
3. Garcia, Carlos, Francisco Molina, Enrique Roca, and Juan M. Lema. “Fuzzy-Based Control of an Anaerobic Reactor Treating Wastewaters Containing Ethanol and Carbohydrates.” Industrial & Engineering Chemistry Research 46, no. 21 (October 1, 2007): 6707–15. https://doi.org/10.1021/ie0617001. Accessed on 31 December 2021
4. Alsina, E.F.; Chica, M.; Trawi ´nski, K.; Regattieri, A. On the use of machine learning methods to predict component reliability from data-driven industrial case studies. Int. J. Adv. Manuf. Technol. 2018, 5, 2419–2433. Accessed on 20 August 2020
5. Ata Allah Nadiri, Sima Shokri, Frank T.-C. Tsai, Asghar Asghari Moghaddam, Prediction of Effluent Quality Parameters of a Wastewater Treatment Plant Using a Supervised Committee Fuzzy Logic Model, Journal of Cleaner Production (2018), doi: 10.1016/j. jclepro.2018.01.139. Accessed on 20 August 2020.
6. E. B. Hassen and A. M. Asmare. Predictive performance modeling of Habesha brewery wastewater treatment plant using artificial neural networks. Chemistry International 5(1) (2019) 87-96. Accessed on 20 August 2020.
7. Wastewater treatment plant performance analysis using artificial intelligence – an ensemble approach- Vahid Nourani, Gozen Elkiran and S. I. Abba . Accessed on 20 August 2020.
8. Use of artificial neural network black-box modeling for the prediction of Wastewater treatment plants performance-ScienceDirect.

   https://www.sciencedirect.com/science/article/pii/S0301479706001228. Accessed on 27 June 2021

9. Han, Honggui, Hongxu Liu, and Junfei Qiao. “Knowledge-Data-Driven Flexible Switching Control for Wastewater Treatment Process.” IEEE Transactions on Control

Systems Technology, 2021, 1–14. https://doi.org/10.1109/TCST.2021.3095849. Accessed on 15 October 2021

10. H. Liu, H. Zhang, Y. Zhang, F. Zhang and M. Huang, "Modeling of Wastewater Treatment Processes Using Dynamic Bayesian Networks Based on Fuzzy PLS," in IEEE Access, vol. 8, pp. 92129-92140, 2020, doi: 10.1109/ACCESS.2020.2995068. Accessed on 1 January 2021
11. T. Yu and Y. Bai, "A Comparative Study of Extreme Learning Machine, Least Squares Support Vector Machine, Back Propagation Neural Network for Outlet Total Phosphorus Prediction," 2018 Prognostics and System Health Management Conference (PHM-Chongqing), 2018, pp. 717-722, doi: 10.1109/PHM-Chongqing.2018.00128. Accessed on 1 January 2021.
12. "Water Quality." https://kaggle.com/adityakadiwal/water-potability. Accessed January 14, 2022.
13. https://www.sciencedirect.com/topics/engineering/confusion-matrix

## APPENDIX

### Sample Code

```
import pandas as pd
import numpy as np

from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier

url="water_potability.csv"
df=pd.read_csv(url)
print(df.shape, df.shape)
print(len(df), len(df))
X = df.drop(['Potability'], axis = 1)
Y = df['Potability']
X_train, X_test, y_train, y_test = train_test_split(X,Y,test_size=0.10)
from sklearn.tree import DecisionTreeClassifier
dt = DecisionTreeClassifier(criterion = 'entropy', random_state = 42)
dt.fit(X_train, y_train)
dt_pred_train = dt.predict(X_train)
```

## Project Presentation for Srishti

SRISHTI 2022 Selection Intimation - Project Id: ..INCS02..... Inbox ×

NISHA JOSEPH CS <nisha.joseph@saintgits.org> Thu, Apr 21, 9:19 PM
to MANNA, me, SIAN, SHERIN, NISHA, POLLY, JEYAKRISHNAN, JINU, DIVYA, HOD

**Greetings from SAINTGITS College of Engineering (Autonomous)**

**Congratulations!!!**

**Your Project titled "…** Artificial Intelligence based Data driven Decision Support System in Waste Water Treatment Plants **…" with project ID………INCS02………….. is selected for the final round of SRISHTI 2022 scheduled on 25th and 26th April.**

**Please confirm your participation by sending a reply e-mail and register using the link (Google form) given below on or before Friday, 22th April 2022 by 10.00 am.**

**Details provided in the form will be used for certification.**

**Click here to register & confirm your participation**

https://forms.gle/HN4mdQPKav3PnSuf8

## Paper Publication for ISBN on behalf of Srishti

Regarding submission of abstract for publication SRISHTI 2022 Inbox ×

SRISHTI 2022 <srishti2022@saintgits.org> Tue, May 10, 2:37 PM
to me, Nisha

Dear Sir/Madam,

Greetings from Saintgits College of Engineering (Autonomous), Kottayam

As mentioned in the SRISHTI 2022 brochure, we consider your completed project's abstract for publication in an ISBN numbered book. If you wish to publish your abstract, kindly follow the instructions given in the abstract template. The abstract template is attached along with this mail and the abstracts submitted in any other formats will be summarily rejected.

1. **Only registered participants with a project ID can submit the extended abstract**
2. Use the given template for making the extended project abstract
3. Never exceed more than **four** pages
4. Submit your abstracts here https://forms.gle/8yuiPrFrAX6PXQi58
5. Plagiarized project abstracts will be summarily rejected
6. **Last Date of submission of Extended Abstract: 26-05-2022, 5 PM**

With Regards

Team SRISHTI 2022

Your Project ID CS328

# Artificial Intelligence based Data-Driven Decision Support System in Wastewater Treatment Plants

Manna Ann Mariyam, Sherin John , Sian Gijo , Winny Anna Varkey
Saintgits College of Engineering, Kottukulam Hills , Pathamuttom, Kottayam , Kerala
nisha.joseph@saintgits.org

**Abstract :**

The Wastewater Treatment Plant (WWTP) plays a crucial role in our society that involves combination of complex processes for treating and removing pollutants from water. Conventionally decisions in the WWTPs are taken by skilled technicians who are highly qualified in their fields. A considerable amount of error can occur when critical decisions are taken by these operators. In order to tackle this and to improve efficiency and accuracy, the proposed system is a data-driven Decision Support System based on Artificial Intelligence that can be used in WWTPs as traditional methods of decision making by human operators are considerably less efficient.

## Introduction

Clean water is one of the most essential natural resources on the earth because it is necessary for life. The elimination of contaminants from wastewater before it reaches aquifers or natural bodies of water is known as wastewater treatment. Water is said to be polluted when it contains enough contaminants to make it inappropriate for a specific use, such as drinking, irrigation etc. Water pollution is mostly caused by the discharge of contaminated wastewater into bodies of water, and wastewater treatment is a key component of water pollution prevention. Water treatment plants (WWTP) are built to accelerate the natural purification process.

Wastewater treatment is critical for the health of a variety of ecosystems. When wastewater is adequately treated, it can be used as a supply of water for a variety of uses. When wastewater is treated properly, the maximum quantity of water can be utilized rather than being wasted.

## Project description

Wastewater is used water that might contain a lot of contaminants, so it is required to be treated before released into the environment or reused for other purposes. Thus wastewater treatment proves to be critical. Traditionally, all decisions in a WWTP are made by experts and technicians, making it vulnerable to a variety of errors. Beginner mistakes with wastewater flow, pumping, inlet and outlet hydraulics, detention time, aeration, and sludge treatment reveal a clear lack of professional exposure. Lack of expertise manifests itself in the frequent copying of foreign solutions and old initiatives, without regard for local circumstances or other significant impacts.

In this field, a lack of systematic specialist education might have serious consequences. If wrong decisions are taken at various stages, this will directly affect the treatment of water. When this ill treatment is released into the environment, it will badly affect the ecosystem and its inhabitants.So the requirement of an