**Project Report On**

**Water quality prediction analysis**

**Submitted by**

**E-ADA-085-Anchuri Yamuna**

**E-ADA-087-Ankam Ravi Varma**

**E-ADA-114-Lokesh Sonwani**

**E-ADA-123-Muddada Anusha**

**E-ADA-135-Prerna Sahu**

**E-ADA-164-Vallepu Prabhu Das**

**Applied Data Analytics - A practical approach**

**Continuing Education Cell**

**National Institute of Technology, Raipur**

**Github -** [**https://github.com/ravi-varma-ankam/water\_quality\_prediction**](https://github.com/ravi-varma-ankam/water_quality_prediction)

**Summary:-**

Water quality analysis is a complex topic due to the different factors that influence it. This concept is inextricably linked to the various purposes for which water is used. Different needs necessitate different standards. There is a lot of study being done on water quality prediction. Water quality is normally determined by a set of physical and chemical parameters that are closely related to the water's intended usage. The acceptable and unacceptable values for each variable must then be established. Water that meets the pre-determined parameters for a specific application is considered appropriate for that application. If the water does not fulfil these requirements, it must be treated before it may be used. So this is how you can analyze the prediction of water quality and train a machine learning model to classify safe and unsafe water for drinking. This work shows the comparative analysis of different machine learning approaches like Logistic regression, Support Vector Machine (SVM), Decision Tree (DT), Random forest classifier, K-Nearest-Neighbours (KNN). From a legal point of view, access to drinking water is one of the fundamental human rights. Many factors affect water quality, it is also one of the major research areas in machine learning.

**Conclusion:-**

In any aquatic system analysis, the modelling water quality parameters are of considerable significance. The traditional modelling methodologies are dependent on datasets that involve large amount of unknown or unspecified input data and generally consist of time-consuming processes. The implementation of Machine Learning (ML) leads to a flexible mathematical structure that has the capability to identify non-linear and complex relationships between input and output data.We must first acquire data for our model data in order to figure out the aspects which would be the most beneficial for forecasting models. We need to conduct data preprocessing in order to rectify any flaws which may appear in our dataset, including such missing values, improperly adjusted data. Then, in order to rate the model, we will split our dataset into two portions: Train and Test. After that, we will utilize our dataset to deploy a machine learning model. In this case, using a prescriptive analysis based on projected values would result in future capabilities to assist decision and policy makers.

We conclude that we predict the water quality with machine learning models. Here we trained the data in various models like logistic regression, k nearest neighbors, support vector machine, random forest classifier and decision tree. Here we get most accuracy in svm(suppport vector machine) as explained above with the mentioned bar graph. There we get accuracy of **0.685976** in svm compared to other models.