

Title: Efficient Water Pliability Prediction for Human Consumption-Exploratory data analysis (EDA) and Multi Algos

Author Names: Tanzina Afroz Rimi; Tanmoy Komer Readoy; Md Tahmid Chowdhury; Sheak Rashed Haider Noori; Narayan Ranjan Chakraborty; Mayen Uddin Mojumdar

Abstract: Water quality assessment is crucial for public health, prompting the gathering and preprocessing of a comprehensive dataset encompassing diverse quality parameters. This study focuses on enhancing the prediction of water suitability for human consumption through the utilization of exploratory data analysis (EDA) and a multi-algorithm approach. Two Kaggle water quality datasets are merged into one, and an additional class of 'Usable but Non Drinkable' is introduced. In the dataset, patterns and anomalies are identified through exploratory data analysis. The performance of several methods, including Decision Trees, Random Forests, SVM, KNN, Gradient Boosting, and a Voting Classifier, is maximized using machine learning and CNN. Each algorithm undergoes extensive training, evaluation, and tuning, and its performance is measured using a variety of metrics. Random Forests achieved the highest accuracy of 85.76%. The paper outlines algorithmic advantages and disadvantages to aid in selecting the best algorithms for predicting water pliability. The ensemble method utilized by the Voting Classifier highlights the advantages of algorithm fusion

Key words: Water quality assessment; Machine learning; Exploratory data analysis (EDA); Decision Trees; Random Forest; Support Vector Machine (SVM)

DOI: [10.1109/ICCCNT61001.2024.10724388](https://doi.org/10.1109/ICCCNT61001.2024.10724388)

<https://ieeexplore.ieee.org/document/10724388>