

Water Quality Prediction Project Report

Project Title: Water Quality Prediction Using Machine Learning

Objective: Predict water quality based on physicochemical parameters to identify safe and unsafe water.

Introduction

Monitoring water quality is essential for public health. Machine learning can predict water quality efficiently using dataset parameters.

Dataset Overview

- **Features:** pH, Hardness, Solids, Chloramines, Sulfate, Conductivity, Organic Carbon, Trihalomethanes, Turbidity, etc.
 - **Target:** Water Quality (Safe / Unsafe)
-

Methodology

1. **Preprocessing:** Handle missing values, scale features, encode targets.
 2. **Models Used:** Logistic Regression, Decision Tree, KNN, SVM, AdaBoost, Random Forest, XGBoost.
 3. **Evaluation:** Accuracy Score
-

Results

Model	Accuracy
SVM	0.695
XGBoost	0.668
Random Forest	0.665
Decision Tree	0.654
AdaBoost	0.634
KNN	0.630
Logistic Regression	0.628

Conclusion

- SVM and XGBoost performed best.
 - Ensemble and non-linear models capture complex patterns in water quality better than linear models.
 - Machine learning aids in timely water quality monitoring.
-

References

Kaggle - Water Quality Prediction