A Machine Learning–Powered Water Quality Prediction Model and Cost-Effective IOT Water Quality Monitoring System

Mohsin, Humzah (School: Valley Stream South High School)

This study presents the design, implementation, and validation of a low-cost autonomous water quality monitoring and forecasting system integrating embedded IoT hardware with an adaptive machine learning pipeline. The system centers on an ESP32-WROOM-32 microcontroller with integrated WiFi, interfacing with waterproof temperature and total dissolved solids sensors via 1-Wire and analog protocols, transmitting data via HTTP POST to a Raspberry Pi server hosting a full-stack web application with real-time visualization, geographic heatmapping, and configurable threshold alerting. Multi-point calibration against laboratory reference standards yielded R-squared of 0.9998 for temperature with mean absolute error of 0.16 degrees Celsius, and R-squared of 0.9987 for TDS with error below 3 percent, demonstrating consumer-grade components achieve near-laboratory accuracy through rigorous validation. An automated alert system achieved 100 percent detection accuracy across 60 trials with mean response latency of 8.3 seconds. A Random Forest regression model incorporating 44 engineered features including cyclic temporal encodings, lag features, and rolling statistical windows was trained on 8,760 hourly observations to forecast temperature and TDS at 7, 14, and 30-day horizons. A continuous retraining mechanism updates model parameters weekly on accumulated site-specific data, enabling adaptive localization without manual intervention. Model performance exceeded R-squared of 0.90 across all horizons with MAPE below 4 percent. Total system cost is under $120, approximately 4 percent of comparable commercial platforms, demonstrating production-grade sensing and forecasting through principled embedded systems engineering at accessible cost.

Awards Won: