

ABSTRAK

Air bersih dan aman sangat penting bagi kesehatan manusia dan diakui sebagai hak dasar. Namun, penurunan kualitas air secara global akibat pencemaran dan pengelolaan air yang tidak memadai menimbulkan risiko serius bagi kesehatan masyarakat, sehingga evaluasi kelayakan air minum menjadi sangat penting. Penelitian ini bertujuan untuk mengklasifikasikan kelayakan air minum secara lebih akurat dengan memanfaatkan model *machine learning*. Penelitian ini menggunakan *Water Potability Dataset* yang terdiri dari 3.276 data dan 10 parameter seperti *pH*, *hardness*, dan *solids*, yang dianalisis untuk memberikan penilaian kualitas air yang komprehensif. Nilai yang hilang pada data ditangani dengan teknik imputasi, sedangkan ketidakseimbangan kelas diatasi dengan metode penyeimbangan data. *Dataset* dibagi menjadi 80% data pelatihan dan 20% data pengujian untuk memastikan validasi model yang andal. Model *Decision Tree* dan *Random Forest* diterapkan karena kemampuannya dalam mengelola *dataset* kompleks dan menghasilkan output yang dapat diinterpretasikan. Model *Random Forest* memberikan hasil yang lebih baik dibandingkan *Decision Tree*, dengan akurasi lebih tinggi dan keandalan yang unggul dalam mengklasifikasikan air sebagai layak atau tidak layak minum. Hasil evaluasi menunjukkan bahwa model ini efektif dalam mengklasifikasikan air, serta memperlihatkan potensi *machine learning* dalam meningkatkan analisis kualitas air. Temuan ini menyediakan kerangka kerja berbasis data yang dapat mendukung pengambil kebijakan, otoritas pengelola air, dan organisasi kesehatan masyarakat dalam menghadapi tantangan kualitas air. Dengan mendukung penilaian yang tepat waktu dan akurat, penelitian ini berkontribusi dalam melindungi kesehatan masyarakat dan mempromosikan praktik pengelolaan air yang berkelanjutan. Penelitian ini juga disusun sebagai bagian dari tugas akhir dan saat ini sedang dalam proses pengajuan publikasi, agar hasilnya dapat memberikan kontribusi lebih luas bagi penelitian akademik maupun penerapan nyata di lapangan.

Kata Kunci: Data Tidak Seimbang, Penanganan Data Hilang, Algoritma Pohon Keputusan, Model Random Forest, Risiko Kontaminasi Air, Penelitian